

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
28 March 2002 (28.03.2002)

PCT

(10) International Publication Number  
**WO 02/24718 A1**

(51) International Patent Classification<sup>7</sup>: **C07H 21/02**,  
21/04, C12N 15/85, C12P 21/06

[IN/US]; 3475 Granada Avenue, #357, Santa Clara, CA  
95051 (US).

(21) International Application Number: PCT/US01/29386

(74) Agents: **LICATA, Jane, Massey et al.**; Licata & Tyrrell  
P.C., 66 E. Main Street, Marlton, NJ 08053 (US).

(22) International Filing Date:

19 September 2001 (19.09.2001)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

60/233,746 19 September 2000 (19.09.2000) US

(63) Related by continuation (CON) or continuation-in-part  
(CIP) to earlier application:

US 60/233,746 (CIP)  
Filed on 19 September 2000 (19.09.2000)

(71) Applicant (for all designated States except US): **DI-  
ADEXUS, INC.** [US/US]; 3303 Octavius Drive, Santa  
Clara, CA 95054 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **SUN, Yongming**  
[US/US]; Apartment 260, 869 S. Winchester Boulevard,  
San Jose, CA 95128 (US). **RECIPON, Herve** [FR/US];  
85 Fortuna Avenue, San Francisco, CA 94115 (US).  
**CAFFERKEY, Robert** [IE/US]; 350 Elan Village Lane,  
Apartment 218, San Jose, CA 95134 (US). **ALI, Shujath**

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,  
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,  
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,  
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,  
MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,  
SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA,  
ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian  
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European  
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,  
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,  
CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,  
TG).

**Published:**

- with international search report
- before the expiration of the time limit for amending the  
claims and to be republished in the event of receipt of  
amendments

For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.

WO 02/24718 A1

(54) Title: COMPOSITIONS AND METHODS RELATING TO PROSTATE SPECIFIC GENES AND PROTEINS

(57) Abstract: The present invention relates to newly identified nucleic acid and polypeptides present in normal and neoplastic fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acid, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating prostate cancer and non-cancerous disease states in prostate, identifying prostate tissue, monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered prostate tissue for treatment and research.

5

10

15

**COMPOSITIONS AND METHODS****RELATING TO PROSTATE SPECIFIC GENES AND PROTEINS**

20

This application claims the benefit of priority from U.S. provisional application Serial No. 60/233,746, filed September 19, 2000.

**FIELD OF THE INVENTION**

25

The present invention relates to newly identified nucleic acids and polypeptides present in normal and neoplastic prostate cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the

30 nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating prostate cancer and

non-cancerous disease states in prostate, identifying prostate tissue and monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered prostate tissue for treatment and research.

5

## BACKGROUND OF THE INVENTION

Prostate cancer is the most prevalent cancer in men and is the second leading cause of death from cancer among males in the United States. AJCC Cancer Staging Handbook 203 (Irvin D. Fleming et al. eds., 5<sup>th</sup> ed. 1998); Walter J. Burdette, Cancer: Etiology, Diagnosis, and Treatment 147 (1998). In 1999, it was estimated that 37,000 men in the United States would die as result of prostate cancer. Elizabeth A. Platz et al., & Edward Giovannucci, *Epidemiology of and Risk Factors for Prostate Cancer, in Management of Prostate Cancer* 21 (Eric A Klein, ed. 2000). Cancer of the prostate typically occurs in older males, with a median age of 74 years for clinical diagnosis. Burdette, *supra* at 147. A man's risk of being diagnosed with invasive prostate cancer in his lifetime is one in six. Platz et al., *supra* at 21.

Although our understanding of the etiology of prostate cancer is incomplete, the results of extensive research in this area point to a combination of age, genetic and environmental/dietary factors. Platz et al., *supra* at 19; Burdette, *supra* at 147; Steven K. Clinton, *Diet and Nutrition in Prostate Cancer Prevention and Therapy, in Prostate Cancer: A Multidisciplinary Guide* 246-269 (Philip W. Kantoff et al. eds. 1997).

Broadly speaking, genetic risk factors predisposing one to prostate cancer include race and a family history of the disease. Platz et al., *supra* at 19, 28-29, 32-34. Aside from these generalities, a deeper understanding of the genetic basis of prostate cancer has remained elusive. Considerable research has been directed to studying the link between prostate cancer, androgens, and androgen regulation, as androgens play a crucial role in

prostate growth and differentiation. Meena Augustus et al., *Molecular Genetics and Markers of Progression, in Management of Prostate Cancer* 59 (Eric A Klein ed. 2000).

While a number of studies have concluded that prostate tumor development is linked to elevated levels of circulating androgen (e.g., testosterone and dihydrotestosterone), the genetic determinants of these levels remain unknown. Platz et al., *supra* at 29-30.

Several studies have explored a possible link between prostate cancer and the androgen receptor (AR) gene, the gene product of which mediates the molecular and cellular effects of testosterone and dihydrotestosterone in tissues responsive to androgens. *Id.* at 30. Differences in the number of certain trinucleotide repeats in exon 1, the region involved in transactivational control, have been of particular interest. Augustus et al., *supra* at 60. For example, these studies have revealed that as the number of CAG repeats decreases the transactivation ability of the gene product increases, as does the risk of prostate cancer. Platz et al., *supra* at 30-31. Other research has focused on the  $\alpha$ -reductase Type 2 gene, the gene which codes for the enzyme that converts testosterone into dihydrotestosterone. *Id.* at 30. Dihydrotestosterone has greater affinity for the AR than testosterone, resulting in increased transactivation of genes responsive to androgens. *Id.* While studies have reported differences among the races in the length of a TA dinucleotide repeat in the 3' untranslated region, no link has been established between the length of that repeat and prostate cancer. *Id.*

Interestingly, while *ras* gene mutations are implicated in numerous other cancers, such mutations appear not to play a significant role in prostate cancer, at least among Caucasian males. Augustus, *supra* at 52.

Environmental/dietary risk factors which may increase the risk of prostate cancer include intake of saturated fat and calcium. Platz et al., *supra* at 19, 25-26. Conversely, intake of selenium, vitamin E and tomato products (which contain the carotenoid lycopene) apparently decrease that risk. *Id.* at 19, 26-28 The impact of physical activity, cigarette smoking, and alcohol consumption on prostate cancer is unclear. Platz et al., *supra* at 23-25.



Periodic screening for prostate cancer is most effectively performed by digital rectal examination (DRE) of the prostate, in conjunction with determination of the serum level of prostate-specific antigen (PSA). Burdette, *supra* at 148. While the merits of such screening are the subject of considerable debate, Jerome P. Richie & Irving D. Kaplan, *Screening for Prostate Cancer: The Horns of a Dilemma*, in Prostate Cancer: A Multidisciplinary Guide 1-10 (Philip W. Kantoff et al. eds. 1997), the American Cancer Society and American Urological Association recommend that both of these tests be performed annually on men 50 years or older with a life expectancy of at least 10 years, and younger men at high risk for prostate cancer. Ian M. Thompson & John Foley, *Screening for Prostate Cancer*, in Management of Prostate Cancer 71 (Eric A Klein ed. 2000). If necessary, these screening methods may be followed by additional tests, including biopsy, ultrasonic imaging, computerized tomography, and magnetic resonance imaging. Christopher A. Haas & Martin I. Resnick, *Trends in Diagnosis, Biopsy, and Imaging*, in Management of Prostate Cancer 89-98 (Eric A Klein ed. 2000); Burdette, *supra* at 148.

Once the diagnosis of prostate cancer has been made, treatment decisions for the individual are typically linked to the stage of prostate cancer present in that individual, as well as his age and overall health. Burdette, *supra* at 151. One preferred classification system for staging prostate cancer was developed by the American Urological Association (AUA). *Id.* at 148. The AUA classification system divides prostate tumors into four broad stages, A to D, which are in turn accompanied by a number of smaller substages. Burdette, *supra* at 152-153; Anthony V. D'Amico et al., *The Staging of Prostate Cancer*, in Prostate Cancer: A Multidisciplinary Guide 41 (Philip W. Kantoff et al. eds. 1997).

Stage A prostate cancer refers to the presence of microscopic cancer within the prostate gland. D'Amico, *supra* at 41. This stage is comprised of two substages: A1, which involves less than four well-differentiated cancer foci within the prostate, and A2, which involves greater than three well-differentiated cancer foci or alternatively, moderately to poorly differentiated foci within the prostate. Burdette, *supra* at 152;

D'Amico, *supra* at 41. Treatment for stage A1 preferentially involves following PSA levels and periodic DRE. Burdette, *supra* at 151. Should PSA levels rise, preferred treatments include radical prostatectomy in patients 70 years of age and younger, external beam radiotherapy for patients between 70 and 80 years of age, and hormone therapy for those over 80 years of age. *Id.*

Stage B prostate cancer is characterized by the presence of a palpable lump within the prostate. Burdette, *supra* at 152-53; D'Amico, *supra* at 41. This stage is comprised of three substages: B1, in which the lump is less than 2 cm and is contained in one lobe of the prostate; B2, in which the lump is greater than 2 cm yet is still contained within one lobe; and B3, in which the lump has spread to both lobes. Burdette, *supra*, at 152-53. For stages B1 and B2, the treatment again involves radical prostatectomy in patients 70 years of age and younger, external beam radiotherapy for patients between 70 and 80 years of age, and hormone therapy for those over 80 years of age. *Id.* at 151. In stage B3, radical prostatectomy is employed if the cancer is well-differentiated and PSA levels are below 15 ng/mL; otherwise, external beam radiation is the chosen treatment option. *Id.*

Stage C prostate cancer involves a substantial cancer mass accompanied by extraprostatic extension. Burdette, *supra* at 153; D'Amico, *supra* at 41. Like stage A prostate cancer, Stage C is comprised of two substages: substage C1, in which the tumor is relatively minimal, with minor prostatic extension, and substage C2, in which the tumor is large and bulky, with major prostatic extension. *Id.* The treatment of choice for both substages is external beam radiation. Burdette, *supra* at 151.

The fourth and final stage of prostate cancer, Stage D, describes the extent to which the cancer has metastasized. Burdette, *supra* at 153; D'Amico, *supra* at 41. This stage is comprised of four substages: (1) D0, in which acid phosphatase levels are persistently high, (2) D1, in which only the pelvic lymph nodes have been invaded, (3) D2, in which the lymph nodes above the aortic bifurcation have been invaded, with or without distant metastasis, and (4) D3, in which the metastasis progresses despite intense

hormonal therapy. *Id.* Treatment at this stage may involve hormonal therapy, chemotherapy, and removal of one or both testes. Burdette, *supra* at 151.

Despite the need for accurate staging of prostate cancer, current staging methodology is limited. The wide variety of biological behavior displayed by neoplasms of the prostate has resulted in considerable difficulty in predicting and assessing the course of prostate cancer. Augustus et al., *supra* at 47. Indeed, despite the fact that most prostate cancer patients have carcinomas that are of intermediate grade and stage, prognosis for these types of carcinomas is highly variable. Andrew A Renshaw & Christopher L. Corless, *Prognostic Features in the Pathology of Prostate Cancer*, in Prostate Cancer: A Multidisciplinary Guide 26 (Philip W. Kantoff et al. eds. 1997). Techniques such as transrectal ultrasound, abdominal and pelvic computerized tomography, and MRI have not been particularly useful in predicting local tumor extension. D'Amico, *supra* at 53 (editors' comment). While the use of serum PSA in combination with the Gleason score is currently the most effective method of staging prostate cancer, *id.*, PSA is of limited predictive value, Augustus et al., *supra* at 47; Renshaw et al., *supra* at 26, and the Gleason score is prone to variability and error, King, C. R. & Long, J. P., *Int'l. J. Cancer* 90(6): 326-30 (2000). As such, the current focus of prostate cancer research has been to obtain biomarkers to help better assess the progression of the disease. Augustus et al., *supra* at 47; Renshaw et al., *supra* at 26; Pettaway, C. A., *Tech. Urol.* 4(1): 35-42 (1998).

Accordingly, there is a great need for more sensitive and accurate methods for predicting whether a person is likely to develop prostate cancer, for diagnosing prostate cancer, for monitoring the progression of the disease, for staging the prostate cancer, for determining whether the prostate cancer has metastasized and for imaging the prostate cancer. There is also a need for better treatment of prostate cancer.

**SUMMARY OF THE INVENTION**

The present invention solves these and other needs in the art by providing nucleic acid molecules, polypeptides and antibodies thereto, variants and derivatives of the nucleic acids and polypeptides, agonists and antagonists that may be used to identify, 5 diagnose, monitor, stage, image and treat prostate cancer and non-cancerous disease states in prostate; identify and monitor prostate tissue; and identify and design agonists and antagonists of polypeptides of the invention. The invention also provides gene therapy, methods for producing transgenic animals and cells, and methods for producing engineered prostate tissue for treatment and research.

10 In one embodiment, the invention provides nucleic acid molecules that are specific to prostate cells, prostate tissue and/or the prostate organ. These prostate specific nucleic acids (PSNAs) may be a naturally-occurring cDNA, genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. If the PSNA is genomic DNA, then the PSNA is a prostate 15 specific gene (PSG). In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to prostate. In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO:23-31. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO:1-22.

20 In another aspect, the invention provides a nucleic acid molecule that selectively hybridizes or exhibits substantial sequence similarity to a nucleic acid molecule encoding a prostate specific protein (PSP), or that selectively hybridizes or exhibits substantial sequence similarity to a PSNA. In another aspect, the invention provides a nucleic acid molecule that is an allelic variant of a nucleic acid molecule encoding a PSP, or that is an 25 allelic variant of a PSNA. A further object of the invention is to provide a nucleic acid molecule that comprises a part of a nucleic acid sequence that encodes a PSP or that comprises a part of a nucleic acid sequence of a PSNA.

A still further object of the invention is to provide a nucleic acid molecule comprising one or more expression control sequences controlling the transcription and/or 30 translation of all or a part of a PSNA. Another object of the invention is to provide a nucleic acid molecule comprising one or more expression control sequences controlling

the transcription and/or translation of a nucleic acid molecule that encodes all or a fragment of a PSP.

Another object of the invention is to provide a vector and/or host cell comprising a nucleic acid molecule of the instant invention. In a preferred embodiment, the nucleic acid molecule encodes all or a fragment of a PSP. In another preferred embodiment, the nucleic acid molecule comprises all or a part of a PSNA. A further object of the invention is to use the host cell comprising the nucleic acid molecule of the invention to recombinantly produce polypeptides of the invention.

Another object of the invention is to provide a polypeptide encoded by a nucleic acid molecule of the invention. In a preferred embodiment, the polypeptide is a PSP. The polypeptide may comprise either a fragment or a full-length protein.

A further aspect of the invention is to provide polypeptides that are mutant proteins (muteins), fusion proteins, homologous proteins and polypeptides encoded by allelic variants of the PSPs provided herein.

Another object of the invention is to provide an antibody that specifically binds to a polypeptide of the instant invention. The invention also provides an antibody that can bind to a mutein, fusion protein, a homologous protein or a polypeptides encoded by allelic variants of the PSPs provided herein.

A further object of the invention is to provide agonists and antagonists of the nucleic acid molecules and polypeptides of the instant invention.

The instant invention provides methods for using the nucleic acid molecules to detect or amplify nucleic acid molecules that have similar or identical nucleic acid sequences compared to the nucleic acid molecules described herein. In a preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying, diagnosing, monitoring, staging, imaging and treating prostate cancer and non-cancerous disease states in prostate. In another preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying and/or monitoring prostate tissue. The nucleic acid molecules of the instant invention may also be used in gene therapy, for producing transgenic animals and cells, and for producing engineered prostate tissue for treatment and research.

The polypeptides and/or antibodies of the instant invention may be used to identify, diagnose, monitor, stage, image and treat prostate cancer and non-cancerous disease states in prostate. The invention provides methods of using the polypeptides of the invention to identify and/or monitor prostate tissue, and to produce engineered  
5 prostate tissue.

The agonists and antagonists of the instant invention may be used to treat prostate cancer and non-cancerous disease states in prostate and to produce engineered prostate tissue.

Another objective of the invention is to provide a computer readable means of  
10 storing the nucleic acid and amino acid sequences of the invention. The records of the computer readable means can be accessed for reading and displaying of sequences for comparison, alignment and ordering of the sequences of the invention to other sequences.

## DETAILED DESCRIPTION OF THE INVENTION

### 15 Definitions and General Techniques

Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by those of ordinary skill in the art. Further, unless otherwise required by context, singular  
20 terms shall include pluralities and plural terms shall include the singular. Generally, nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well-known and commonly used in the art. The methods and techniques of the present invention are generally performed  
25 according to conventional methods well-known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. *See, e.g.,* Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press (1989) and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 3d ed., Cold Spring Harbor  
30 Press (2001); Ausubel *et al.*, Current Protocols in Molecular Biology, Greene Publishing Associates (1992, and Supplements to 2000); Ausubel *et al.*, Short Protocols in

Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology – 4<sup>th</sup> Ed., Wiley & Sons (1999); Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1990); and Harlow and Lane, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1999); each of  
5 which is incorporated herein by reference in its entirety.

Enzymatic reactions and purification techniques are performed according to manufacturer's specifications, as commonly accomplished in the art or as described herein. The nomenclatures used in connection with, and the laboratory procedures and techniques of, analytical chemistry, synthetic organic chemistry, and medicinal and  
10 pharmaceutical chemistry described herein are those well-known and commonly used in the art. Standard techniques are used for chemical syntheses, chemical analyses, pharmaceutical preparation, formulation, and delivery, and treatment of patients.

The following terms, unless otherwise indicated, shall be understood to have the following meanings:

15 A "nucleic acid molecule" of this invention refers to a polymeric form of nucleotides and includes both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. A nucleotide refers to a ribonucleotide, deoxynucleotide or a modified form of either type of nucleotide. A "nucleic acid molecule" as used herein is synonymous with "nucleic acid" and  
20 "polynucleotide." The term "nucleic acid molecule" usually refers to a molecule of at least 10 bases in length, unless otherwise specified. The term includes single and double stranded forms of DNA. In addition, a polynucleotide may include either or both naturally-occurring and modified nucleotides linked together by naturally-occurring and/or non-naturally occurring nucleotide linkages.

25 The nucleic acid molecules may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates,  
30 phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (*e.g.*, phosphorothioates, phosphorodithioates, etc.), pendent moieties (*e.g.*, polypeptides),

intercalators (*e.g.*, acridine, psoralen, etc.), chelators, alkylators, and modified linkages (*e.g.*, alpha anomeric nucleic acids, etc.) The term “nucleic acid molecule” also includes any topological conformation, including single-stranded, double-stranded, partially duplexed, triplexed, hairpinned, circular and padlocked conformations. Also included  
5 are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

A “gene” is defined as a nucleic acid molecule that comprises a nucleic acid  
10 sequence that encodes a polypeptide and the expression control sequences that surround the nucleic acid sequence that encodes the polypeptide. For instance, a gene may comprise a promoter, one or more enhancers, a nucleic acid sequence that encodes a polypeptide, downstream regulatory sequences and, possibly, other nucleic acid sequences involved in regulation of the expression of an RNA. As is well-known in the  
15 art, eukaryotic genes usually contain both exons and introns. The term “exon” refers to a nucleic acid sequence found in genomic DNA that is bioinformatically predicted and/or experimentally confirmed to contribute contiguous sequence to a mature mRNA transcript. The term “intron” refers to a nucleic acid sequence found in genomic DNA that is predicted and/or confirmed to not contribute to a mature mRNA transcript, but  
20 rather to be “spliced out” during processing of the transcript.

A nucleic acid molecule or polypeptide is “derived” from a particular species if the nucleic acid molecule or polypeptide has been isolated from the particular species, or if the nucleic acid molecule or polypeptide is homologous to a nucleic acid molecule or polypeptide isolated from a particular species.

25 An “isolated” or “substantially pure” nucleic acid or polynucleotide (*e.g.*, an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, *e.g.*, ribosomes, polymerases, or genomic sequences with which it is naturally associated. The term embraces a nucleic acid or polynucleotide that (1) has been  
30 removed from its naturally occurring environment, (2) is not associated with all or a portion of a polynucleotide in which the “isolated polynucleotide” is found in nature, (3)



is operatively linked to a polynucleotide which it is not linked to in nature, (4) does not occur in nature as part of a larger sequence or (5) includes nucleotides or internucleoside bonds that are not found in nature. The term "isolated" or "substantially pure" also can be used in reference to recombinant or cloned DNA isolates, chemically synthesized polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems. The term "isolated nucleic acid molecule" includes nucleic acid molecules that are integrated into a host cell chromosome at a heterologous site, recombinant fusions of a native fragment to a heterologous sequence, recombinant vectors present as episomes or as integrated into a host cell chromosome.

10 A "part" of a nucleic acid molecule refers to a nucleic acid molecule that comprises a partial contiguous sequence of at least 10 bases of the reference nucleic acid molecule. Preferably, a part comprises at least 15 to 20 bases of a reference nucleic acid molecule. In theory, a nucleic acid sequence of 17 nucleotides is of sufficient length to occur at random less frequently than once in the three gigabase human genome, and thus  
15 to provide a nucleic acid probe that can uniquely identify the reference sequence in a nucleic acid mixture of genomic complexity. A preferred part is one that comprises a nucleic acid sequence that can encode at least 6 contiguous amino acid sequences (fragments of at least 18 nucleotides) because they are useful in directing the expression or synthesis of peptides that are useful in mapping the epitopes of the polypeptide encoded by the reference nucleic acid. *See, e.g., Geysen et al., Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1984); and United States Patent Nos. 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. A part may also comprise at least 25, 30, 35 or 40 nucleotides of a reference nucleic acid molecule, or at least 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides of a  
25 reference nucleic acid molecule. A part of a nucleic acid molecule may comprise no other nucleic acid sequences. Alternatively, a part of a nucleic acid may comprise other nucleic acid sequences from other nucleic acid molecules.

The term "oligonucleotide" refers to a nucleic acid molecule generally comprising a length of 200 bases or fewer. The term often refers to single-stranded  
30 deoxyribonucleotides, but it can refer as well to single- or double-stranded ribonucleotides, RNA:DNA hybrids and double-stranded DNAs, among others.

Preferably, oligonucleotides are 10 to 60 bases in length and most preferably 12, 13, 14, 15, 16, 17, 18, 19 or 20 bases in length. Other preferred oligonucleotides are 25, 30, 35, 40, 45, 50, 55 or 60 bases in length. Oligonucleotides may be single-stranded, *e.g.* for use as probes or primers, or may be double-stranded, *e.g.* for use in the construction of a mutant gene. Oligonucleotides of the invention can be either sense or antisense oligonucleotides. An oligonucleotide can be derivatized or modified as discussed above for nucleic acid molecules.

Oligonucleotides, such as single-stranded DNA probe oligonucleotides, often are synthesized by chemical methods, such as those implemented on automated oligonucleotide synthesizers. However, oligonucleotides can be made by a variety of other methods, including in vitro recombinant DNA-mediated techniques and by expression of DNAs in cells and organisms. Initially, chemically synthesized DNAs typically are obtained without a 5' phosphate. The 5' ends of such oligonucleotides are not substrates for phosphodiester bond formation by ligation reactions that employ DNA ligases typically used to form recombinant DNA molecules. Where ligation of such oligonucleotides is desired, a phosphate can be added by standard techniques, such as those that employ a kinase and ATP. The 3' end of a chemically synthesized oligonucleotide generally has a free hydroxyl group and, in the presence of a ligase, such as T4 DNA ligase, readily will form a phosphodiester bond with a 5' phosphate of another polynucleotide, such as another oligonucleotide. As is well-known, this reaction can be prevented selectively, where desired, by removing the 5' phosphates of the other polynucleotide(s) prior to ligation.

The term "naturally-occurring nucleotide" referred to herein includes naturally-occurring deoxyribonucleotides and ribonucleotides. The term "modified nucleotides" referred to herein includes nucleotides with modified or substituted sugar groups and the like. The term "nucleotide linkages" referred to herein includes nucleotides linkages such as phosphorothioate, phosphorodithioate, phosphoroselenoate, phosphorodiselenoate, phosphoroanilothioate, phosphoraniladate, phosphoroamidate, and the like. See *e.g.*, LaPlanche *et al.* *Nucl. Acids Res.* 14:9081-9093 (1986); Stein *et al.* *Nucl. Acids Res.* 16:3209-3221 (1988); Zon *et al.* *Anti-Cancer Drug Design* 6:539-568 (1991); Zon *et al.*, in Eckstein (ed.) Oligonucleotides and Analogues: A Practical

Approach, pp. 87-108, Oxford University Press (1991); United States Patent No. 5,151,510; Uhlmann and Peyman *Chemical Reviews* 90:543 (1990), the disclosures of which are hereby incorporated by reference.

Unless specified otherwise, the left hand end of a polynucleotide sequence in sense orientation is the 5' end and the right hand end of the sequence is the 3' end. In addition, the left hand direction of a polynucleotide sequence in sense orientation is referred to as the 5' direction, while the right hand direction of the polynucleotide sequence is referred to as the 3' direction. Further, unless otherwise indicated, each nucleotide sequence is set forth herein as a sequence of deoxyribonucleotides. It is intended, however, that the given sequence be interpreted as would be appropriate to the polynucleotide composition: for example, if the isolated nucleic acid is composed of RNA, the given sequence intends ribonucleotides, with uridine substituted for thymidine.

The term "allelic variant" refers to one of two or more alternative naturally-occurring forms of a gene, wherein each gene possesses a unique nucleotide sequence. In a preferred embodiment, different alleles of a given gene have similar or identical biological properties.

The term "percent sequence identity" in the context of nucleic acid sequences refers to the residues in two sequences which are the same when aligned for maximum correspondence. The length of sequence identity comparison may be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides. There are a number of different algorithms known in the art which can be used to measure nucleotide sequence identity. For instance, polynucleotide sequences can be compared using FASTA, Gap or Bestfit, which are programs in Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. FASTA, which includes, *e.g.*, the programs FASTA2 and FASTA3, provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson, *Methods Enzymol.* 183: 63-98 (1990); Pearson, *Methods Mol. Biol.* 132: 185-219 (2000); Pearson, *Methods Enzymol.* 266: 227-258 (1996); Pearson, *J. Mol. Biol.* 276: 71-84 (1998); herein incorporated by reference). Unless otherwise specified, default parameters

for a particular program or algorithm are used. For instance, percent sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) or using Gap with its default parameters as provided in GCG Version 6.1, herein incorporated by  
5 reference.

A reference to a nucleic acid sequence encompasses its complement unless otherwise specified. Thus, a reference to a nucleic acid molecule having a particular sequence should be understood to encompass its complementary strand, with its complementary sequence. The complementary strand is also useful, *e.g.*, for antisense  
10 therapy, hybridization probes and PCR primers.

In the molecular biology art, researchers use the terms "percent sequence identity", "percent sequence similarity" and "percent sequence homology" interchangeably. In this application, these terms shall have the same meaning with respect to nucleic acid sequences only.

15 The term "substantial similarity" or "substantial sequence similarity," when referring to a nucleic acid or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 50%, more preferably 60% of the nucleotide bases, usually at least about 70%, more usually at least  
20 about 80%, preferably at least about 90%, and more preferably at least about 95-98% of the nucleotide bases, as measured by any well-known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.

Alternatively, substantial similarity exists when a nucleic acid or fragment thereof hybridizes to another nucleic acid, to a strand of another nucleic acid, or to the  
25 complementary strand thereof, under selective hybridization conditions. Typically, selective hybridization will occur when there is at least about 55% sequence identity — preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90% — over a stretch of at least about 14 nucleotides, more preferably at least  
30 17 nucleotides, even more preferably at least 20, 25, 30, 35, 40, 50, 60, 70, 80, 90 or 100 nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species, length of the complementary regions, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. "Stringent hybridization conditions" and "stringent wash conditions" in the context of nucleic acid hybridization experiments depend upon a number of different physical parameters. The most important parameters include temperature of hybridization, base composition of the nucleic acids, salt concentration and length of the nucleic acid. One having ordinary skill in the art knows how to vary these parameters to achieve a particular stringency of hybridization. In general, "stringent hybridization" is performed at about 25°C below the thermal melting point ( $T_m$ ) for the specific DNA hybrid under a particular set of conditions. "Stringent washing" is performed at temperatures about 5°C lower than the  $T_m$  for the specific DNA hybrid under a particular set of conditions. The  $T_m$  is the temperature at which 50% of the target sequence hybridizes to a perfectly matched probe. See Sambrook (1989), *supra*, p. 9.51, hereby incorporated by reference.

The  $T_m$  for a particular DNA-DNA hybrid can be estimated by the formula:

$$T_m = 81.5^\circ\text{C} + 16.6 (\log_{10}[\text{Na}^+]) + 0.41 (\text{fraction G} + \text{C}) -$$

$$0.63 (\% \text{ formamide}) - (600/l) \text{ where } l \text{ is the length of the hybrid in base pairs.}$$

The  $T_m$  for a particular RNA-RNA hybrid can be estimated by the formula:

$$T_m = 79.8^\circ\text{C} + 18.5 (\log_{10}[\text{Na}^+]) + 0.58 (\text{fraction G} + \text{C}) +$$

$$11.8 (\text{fraction G} + \text{C})^2 - 0.35 (\% \text{ formamide}) - (820/l).$$

The  $T_m$  for a particular RNA-DNA hybrid can be estimated by the formula:

$$T_m = 79.8^\circ\text{C} + 18.5 (\log_{10}[\text{Na}^+]) + 0.58 (\text{fraction G} + \text{C}) +$$

$$11.8 (\text{fraction G} + \text{C})^2 - 0.50 (\% \text{ formamide}) - (820/l).$$

In general, the  $T_m$  decreases by 1-1.5°C for each 1% of mismatch between two nucleic acid sequences. Thus, one having ordinary skill in the art can alter hybridization and/or washing conditions to obtain sequences that have higher or lower degrees of sequence identity to the target nucleic acid. For instance, to obtain hybridizing nucleic acids that contain up to 10% mismatch from the target nucleic acid sequence, 10-15°C would be subtracted from the calculated  $T_m$  of a perfectly matched hybrid, and then the

hybridization and washing temperatures adjusted accordingly. Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well-known in the art.

5           An example of stringent hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 50% formamide/6X SSC at 42°C for at least ten hours and preferably overnight (approximately 16 hours). Another example of stringent hybridization conditions is 6X SSC at 68°C without  
10       formamide for at least ten hours and preferably overnight. An example of moderate stringency hybridization conditions is 6X SSC at 55°C without formamide for at least ten hours and preferably overnight. An example of low stringency hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or northern blot or for screening a  
15       library is 6X SSC at 42°C for at least ten hours. Hybridization conditions to identify nucleic acid sequences that are similar but not identical can be identified by experimentally changing the hybridization temperature from 68°C to 42°C while keeping the salt concentration constant (6X SSC), or keeping the hybridization temperature and salt concentration constant (*e.g.* 42°C and 6X SSC) and varying the formamide  
20       concentration from 50% to 0%. Hybridization buffers may also include blocking agents to lower background. These agents are well-known in the art. *See* Sambrook *et al.* (1989), *supra*, pages 8.46 and 9.46-9.58, herein incorporated by reference. *See also* Ausubel (1992), *supra*, Ausubel (1999), *supra*, and Sambrook (2001), *supra*.

          Wash conditions also can be altered to change stringency conditions. An example  
25       of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see* Sambrook (1989), *supra*, for SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove excess probe. An exemplary medium stringency wash for duplex DNA of more than 100 base pairs is 1x SSC at 45°C for 15 minutes. An exemplary low stringency wash for such a duplex is 4x SSC at 40°C for 15 minutes. In  
30       general, signal-to-noise ratio of 2x or higher than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

As defined herein, nucleic acids that do not hybridize to each other under stringent conditions are still substantially similar to one another if they encode polypeptides that are substantially identical to each other. This occurs, for example, when a nucleic acid is created synthetically or recombinantly using a high codon degeneracy as permitted by the redundancy of the genetic code.

Hybridization conditions for nucleic acid molecules that are shorter than 100 nucleotides in length (*e.g.*, for oligonucleotide probes) may be calculated by the formula:

$$T_m = 81.5^{\circ}\text{C} + 16.6(\log_{10}[\text{Na}^+]) + 0.41(\text{fraction G+C}) - (600/N),$$
 wherein N is change length and the  $[\text{Na}^+]$  is 1 M or less. *See* Sambrook (1989), *supra*, p. 11.46. For hybridization of probes shorter than 100 nucleotides, hybridization is usually performed under stringent conditions (5-10°C below the  $T_m$ ) using high concentrations (0.1-1.0 pmol/ml) of probe. *Id.* at p. 11.45. Determination of hybridization using mismatched probes, pools of degenerate probes or "guessmers," as well as hybridization solutions and methods for empirically determining hybridization conditions are well-known in the art. *See, e.g.*, Ausubel (1999), *supra*; Sambrook (1989), *supra*, pp. 11.45-11.57.

The term "digestion" or "digestion of DNA" refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes referred to herein are commercially available and their reaction conditions, cofactors and other requirements for use are known and routine to the skilled artisan. For analytical purposes, typically, 1  $\mu\text{g}$  of plasmid or DNA fragment is digested with about 2 units of enzyme in about 20  $\mu\text{l}$  of reaction buffer. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50  $\mu\text{g}$  of DNA are digested with 20 to 250 units of enzyme in proportionately larger volumes. Appropriate buffers and substrate amounts for particular restriction enzymes are described in standard laboratory manuals, such as those referenced below, and they are specified by commercial suppliers. Incubation times of about 1 hour at 37°C are ordinarily used, but conditions may vary in accordance with standard procedures, the supplier's instructions and the particulars of the reaction. After digestion, reactions may be analyzed, and fragments may be purified by electrophoresis through an agarose or polyacrylamide gel, using well-known methods that are routine for those skilled in the art.

The term "ligation" refers to the process of forming phosphodiester bonds between two or more polynucleotides, which most often are double stranded DNAs. Techniques for ligation are well-known to the art and protocols for ligation are described in standard laboratory manuals and references, such as, *e.g.*, Sambrook (1989), *supra*.

5       A genome-derived "single exon probes," are probes that comprise at least part of an exon ("reference exon") and can hybridize detectably under high stringency conditions to transcript-derived nucleic acids that include the reference exon but do not hybridize detectably under high stringency conditions to nucleic acids that lack the reference exon. Single exon probes typically further comprise, contiguous to a first end  
10 of the exon portion, a first intronic and/or intergenic sequence that is identically contiguous to the exon in the genome, and may contain a second intronic and/or intergenic sequence that is identically contiguous to the exon in the genome. The minimum length of genome-derived single exon probes is defined by the requirement that the exonic portion be of sufficient length to hybridize under high stringency  
15 conditions to transcript-derived nucleic acids, as discussed above. The maximum length of genome-derived single exon probes is defined by the requirement that the probes contain portions of no more than one exon. The single exon probes may contain priming sequences not found in contiguity with the rest of the probe sequence in the genome, which priming sequences are useful for PCR and other amplification-based technologies.

20       The term "microarray" or "nucleic acid microarray" refers to a substrate-bound collection of plural nucleic acids, hybridization to each of the plurality of bound nucleic acids being separately detectable. The substrate can be solid or porous, planar or non-planar, unitary or distributed. A microarray or nucleic acid microarray include all the devices so called in Schena (ed.), DNA Microarrays: A Practical Approach (Practical  
25 Approach Series), Oxford University Press (1999); *Nature Genet.* 21(1)(suppl.):1 - 60 (1999); Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing Company/BioTechniques Books Division (2000). These microarrays include substrate-bound collections of plural nucleic acids in which the plurality of nucleic acids are disposed on a plurality of beads, rather than on a unitary planar substrate, as is described,  
30 *inter alia*, in Brenner *et al.*, *Proc. Natl. Acad. Sci. USA* 97(4):1665-1670 (2000).



The term "mutated" when applied to nucleic acid sequences means that nucleotides in a nucleic acid sequence may be inserted, deleted or changed compared to a reference nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus.

5 In addition, one or more alterations may be made at any number of loci within a nucleic acid sequence. In a preferred embodiment, the nucleic acid sequence is the wild type nucleic acid sequence encoding a PSP or is a PSNA. The nucleic acid sequence may be mutated by any method known in the art including those mutagenesis techniques described *infra*.

10 The term "error-prone PCR" refers to a process for performing PCR under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. *See, e.g., Leung et al., Technique 1: 11-15 (1989) and Caldwell et al., PCR Methods Applic. 2: 28-33 (1992).*

15 The term "oligonucleotide-directed mutagenesis" refers to a process which enables the generation of site-specific mutations in any cloned DNA segment of interest. *See, e.g., Reidhaar-Olson et al., Science 241: 53-57 (1988).*

The term "assembly PCR" refers to a process which involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR  
20 reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction.

The term "sexual PCR mutagenesis" or "DNA shuffling" refers to a method of error-prone PCR coupled with forced homologous recombination between DNA molecules of different but highly related DNA sequence *in vitro*, caused by random  
25 fragmentation of the DNA molecule based on sequence similarity, followed by fixation of the crossover by primer extension in an error-prone PCR reaction. *See, e.g., Stemmer, Proc. Natl. Acad. Sci. U.S.A. 91: 10747-10751 (1994).* DNA shuffling can be carried out between several related genes ("Family shuffling").

The term "*in vivo* mutagenesis" refers to a process of generating random  
30 mutations in any cloned DNA of interest which involves the propagation of the DNA in a strain of bacteria such as *E. coli* that carries mutations in one or more of the DNA repair

pathways. These “mutator” strains have a higher random mutation rate than that of a wild-type parent. Propagating the DNA in a mutator strain will eventually generate random mutations within the DNA.

5 The term “cassette mutagenesis” refers to any process for replacing a small region of a double-stranded DNA molecule with a synthetic oligonucleotide “cassette” that differs from the native sequence. The oligonucleotide often contains completely and/or partially randomized native sequence.

10 The term “recursive ensemble mutagenesis” refers to an algorithm for protein engineering (protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. *See, e.g., Arkin et al., Proc. Natl. Acad. Sci. U.S.A.* 89: 7811-7815 (1992).

15 The term “exponential ensemble mutagenesis” refers to a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein small groups of residues are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. *See, e.g., Delegrave et al., Biotechnology Research* 11: 1548-1552 (1993); Arnold, *Current Opinion in Biotechnology* 4: 450-455 (1993). Each of the references mentioned above are hereby incorporated by reference in  
20 its entirety.

“Operatively linked” expression control sequences refers to a linkage in which the expression control sequence is contiguous with the gene of interest to control the gene of interest, as well as expression control sequences that act in *trans* or at a distance to control the gene of interest.

25 The term “expression control sequence” as used herein refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operatively linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination,  
30 promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that

enhance translation efficiency (*e.g.*, ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, and transcription termination sequence. The term "control sequences" is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

The term "vector," as used herein, is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments may be ligated. Other vectors include cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Viral vectors that infect bacterial cells are referred to as bacteriophages. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" (or simply, "expression vectors"). In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" may be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include other forms of expression vectors that serve equivalent functions.

The term "recombinant host cell" (or simply "host cell"), as used herein, is intended to refer to a cell into which a recombinant expression vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences,

such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein.

As used herein, the phrase "open reading frame" and the equivalent acronym "ORF" refer to that portion of a transcript-derived nucleic acid that can be translated in its entirety into a sequence of contiguous amino acids. As so defined, an ORF has length, measured in nucleotides, exactly divisible by 3. As so defined, an ORF need not encode the entirety of a natural protein.

As used herein, the phrase "ORF-encoded peptide" refers to the predicted or actual translation of an ORF.

As used herein, the phrase "degenerate variant" of a reference nucleic acid sequence intends all nucleic acid sequences that can be directly translated, using the standard genetic code, to provide an amino acid sequence identical to that translated from the reference nucleic acid sequence.

The term "polypeptide" encompasses both naturally-occurring and non-naturally-occurring proteins and polypeptides, polypeptide fragments and polypeptide mutants, derivatives and analogs. A polypeptide may be monomeric or polymeric. Further, a polypeptide may comprise a number of different modules within a single polypeptide each of which has one or more distinct activities. A preferred polypeptide in accordance with the invention comprises a PSP encoded by a nucleic acid molecule of the instant invention, as well as a fragment, mutant, analog and derivative thereof.

The term "isolated protein" or "isolated polypeptide" is a protein or polypeptide that by virtue of its origin or source of derivation (1) is not associated with naturally associated components that accompany it in its native state, (2) is free of other proteins from the same species (3) is expressed by a cell from a different species, or (4) does not occur in nature. Thus, a polypeptide that is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be "isolated" from its naturally associated components. A polypeptide or protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well-known in the art.

A protein or polypeptide is "substantially pure," "substantially homogeneous" or "substantially purified" when at least about 60% to 75% of a sample exhibits a single

species of polypeptide. The polypeptide or protein may be monomeric or multimeric. A substantially pure polypeptide or protein will typically comprise about 50%, 60%, 70%, 80% or 90% W/W of a protein sample, more usually about 95%, and preferably will be over 99% pure. Protein purity or homogeneity may be indicated by a number of means well-known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualizing a single polypeptide band upon staining the gel with a stain well-known in the art. For certain purposes, higher resolution may be provided by using HPLC or other means well-known in the art for purification.

The term "polypeptide fragment" as used herein refers to a polypeptide that has an amino-terminal and/or carboxy-terminal deletion compared to a full-length polypeptide. In a preferred embodiment, the polypeptide fragment is a contiguous sequence in which the amino acid sequence of the fragment is identical to the corresponding positions in the naturally-occurring sequence. Fragments typically are at least 5, 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or 18 amino acids long, more preferably at least 20 amino acids long, more preferably at least 25, 30, 35, 40 or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even more preferably at least 70 amino acids long.

A "derivative" refers to polypeptides or fragments thereof that are substantially similar in primary structural sequence but which include, *e.g.*, *in vivo* or *in vitro* chemical and biochemical modifications that are not found in the native polypeptide. Such modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. Other modification include, *e.g.*, labeling with radionuclides, and various enzymatic modifications, as will be readily appreciated by

those skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well-known in the art, and include radioactive isotopes such as  $^{125}\text{I}$ ,  $^{32}\text{P}$ ,  $^{35}\text{S}$ , and  $^3\text{H}$ , ligands which bind to labeled antiligands (*e.g.*, antibodies), fluorophores, chemiluminescent agents, enzymes, and  
5 antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods for labeling polypeptides are well-known in the art. *See* Ausubel (1992), *supra*; Ausubel (1999), *supra*, herein incorporated by reference.

10 The term "fusion protein" refers to polypeptides comprising polypeptides or fragments coupled to heterologous amino acid sequences. Fusion proteins are useful because they can be constructed to contain two or more desired functional elements from two or more different proteins. A fusion protein comprises at least 10 contiguous amino acids from a polypeptide of interest, more preferably at least 20 or 30 amino acids, even  
15 more preferably at least 40, 50 or 60 amino acids, yet more preferably at least 75, 100 or 125 amino acids. Fusion proteins can be produced recombinantly by constructing a nucleic acid sequence which encodes the polypeptide or a fragment thereof in frame with a nucleic acid sequence encoding a different protein or peptide and then expressing the fusion protein. Alternatively, a fusion protein can be produced chemically by  
20 crosslinking the polypeptide or a fragment thereof to another protein.

The term "analog" refers to both polypeptide analogs and non-peptide analogs. The term "polypeptide analog" as used herein refers to a polypeptide that is comprised of a segment of at least 25 amino acids that has substantial identity to a portion of an amino acid sequence but which contains non-natural amino acids or non-natural inter-residue  
25 bonds. In a preferred embodiment, the analog has the same or similar biological activity as the native polypeptide. Typically, polypeptide analogs comprise a conservative amino acid substitution (or insertion or deletion) with respect to the naturally-occurring sequence. Analogs typically are at least 20 amino acids long, preferably at least 50 amino acids long or longer, and can often be as long as a full-length naturally-occurring  
30 polypeptide.

The term "non-peptide analog" refers to a compound with properties that are analogous to those of a reference polypeptide. A non-peptide compound may also be termed a "peptide mimetic" or a "peptidomimetic." Such compounds are often developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to useful peptides may be used to produce an equivalent effect. Generally, peptidomimetics are structurally similar to a paradigm polypeptide (*i.e.*, a polypeptide that has a desired biochemical property or pharmacological activity), but have one or more peptide linkages optionally replaced by a linkage selected from the group consisting of: --CH<sub>2</sub>NH--, --CH<sub>2</sub>S--, --CH<sub>2</sub>-CH<sub>2</sub>--, --CH=CH--(*cis* and *trans*), --COCH<sub>2</sub>--, --CH(OH)CH<sub>2</sub>--, and --CH<sub>2</sub>SO--, by methods well-known in the art. Systematic substitution of one or more amino acids of a consensus sequence with a D-amino acid of the same type (*e.g.*, D-lysine in place of L-lysine) may also be used to generate more stable peptides. In addition, constrained peptides comprising a consensus sequence or a substantially identical consensus sequence variation may be generated by methods known in the art (Rizo *et al.*, *Ann. Rev. Biochem.* 61:387-418 (1992), incorporated herein by reference). For example, one may add internal Sistine residues capable of forming intramolecular disulfide bridges which cyclize the peptide.

A "polypeptide mutant" or "mutein" refers to a polypeptide whose sequence contains substitutions, insertions or deletions of one or more amino acids compared to the amino acid sequence of a native or wild-type protein. A mutein may have one or more amino acid point substitutions, in which a single amino acid at a position has been changed to another amino acid, one or more insertions and/or deletions, in which one or more amino acids are inserted or deleted, respectively, in the sequence of the naturally-occurring protein, and/or truncations of the amino acid sequence at either or both the amino or carboxy termini. Further, a mutein may have the same or different biological activity as the naturally-occurring protein. For instance, a mutein may have an increased or decreased biological activity. A mutein has at least 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are muteins having 80%, 85% or 90% sequence similarity to the wild type protein. In an even more preferred embodiment, a mutein

exhibits 95% sequence identity, even more preferably 97%, even more preferably 98% and even more preferably 99%. Sequence similarity may be measured by any common sequence analysis algorithm, such as Gap or Bestfit.

Preferred amino acid substitutions are those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter binding affinity or enzymatic activity, and (5) confer or modify other physicochemical or functional properties of such analogs. For example, single or multiple amino acid substitutions (preferably conservative amino acid substitutions) may be made in the naturally-occurring sequence (preferably in the portion of the polypeptide outside the domain(s) forming intermolecular contacts. In a preferred embodiment, the amino acid substitutions are moderately conservative substitutions or conservative substitutions. In a more preferred embodiment, the amino acid substitutions are conservative substitutions. A conservative amino acid substitution should not substantially change the structural characteristics of the parent sequence (*e.g.*, a replacement amino acid should not tend to disrupt a helix that occurs in the parent sequence, or disrupt other types of secondary structure that characterizes the parent sequence). Examples of art-recognized polypeptide secondary and tertiary structures are described in Creighton (ed.), Proteins, Structures and Molecular Principles, W. H. Freeman and Company (1984); Branden *et al.* (ed.), Introduction to Protein Structure, Garland Publishing (1991); Thornton *et al.*, *Nature* 354:105-106 (1991), each of which are incorporated herein by reference.

As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage. See Golub *et al.* (eds.), Immunology - A Synthesis 2<sup>nd</sup> Ed., Sinauer Associates (1991), which is incorporated herein by reference. Stereoisomers (*e.g.*, D-amino acids) of the twenty conventional amino acids, unnatural amino acids such as  $\alpha$ -,  $\alpha$ -disubstituted amino acids, N-alkyl amino acids, and other unconventional amino acids may also be suitable components for polypeptides of the present invention. Examples of unconventional amino acids include: 4-hydroxyproline,  $\gamma$ -carboxyglutamate,  $\epsilon$ -N,N,N-trimethyllysine,  $\epsilon$ -N-acetyllysine, O-phosphoserine, N-acetylserine, N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, s-N-methylarginine, and other similar amino acids and imino acids (*e.g.*, 4-hydroxyproline). In the polypeptide notation



used herein, the lefthand direction is the amino terminal direction and the right hand direction is the carboxy-terminal direction, in accordance with standard usage and convention.

A protein has "homology" or is "homologous" to a protein from another organism  
5 if the encoded amino acid sequence of the protein has a similar sequence to the encoded amino acid sequence of a protein of a different organism and has a similar biological activity or function. Alternatively, a protein may have homology or be homologous to another protein if the two proteins have similar amino acid sequences and have similar biological activities or functions. Although two proteins are said to be "homologous,"  
10 this does not imply that there is necessarily an evolutionary relationship between the proteins. Instead, the term "homologous" is defined to mean that the two proteins have similar amino acid sequences and similar biological activities or functions. In a preferred embodiment, a homologous protein is one that exhibits 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence  
15 similarity. Even more preferred are homologous proteins that exhibit 80%, 85% or 90% sequence similarity to the wild type protein. In a yet more preferred embodiment, a homologous protein exhibits 95%, 97%, 98% or 99% sequence similarity.

When "sequence similarity" is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino  
20 acid substitutions. In a preferred embodiment, a polypeptide that has "sequence similarity" comprises conservative or moderately conservative amino acid substitutions. A "conservative amino acid substitution" is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (*e.g.*, charge or hydrophobicity). In general, a conservative amino  
25 acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of similarity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well-known to those of skill in the art. *See, e.g., Pearson, Methods Mol.*  
30 *Biol.* 24: 307-31 (1994), herein incorporated by reference.

For instance, the following six groups each contain amino acids that are conservative substitutions for one another:

- 1) Serine (S), Threonine (T);
- 5 2) Aspartic Acid (D), Glutamic Acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);
- 5) Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

10 Alternatively, a conservative replacement is any change having a positive value in the PAM250 log-likelihood matrix disclosed in Gonnet *et al.*, *Science* 256: 1443-45 (1992), herein incorporated by reference. A "moderately conservative" replacement is any change having a nonnegative value in the PAM250 log-likelihood matrix.

Sequence similarity for polypeptides, which is also referred to as sequence  
 15 identity, is typically measured using sequence analysis software. Protein analysis software matches similar sequences using measures of similarity assigned to various substitutions, deletions and other modifications, including conservative amino acid substitutions. For instance, GCG contains programs such as "Gap" and "Bestfit" which can be used with default parameters to determine sequence homology or sequence  
 20 identity between closely related polypeptides, such as homologous polypeptides from different species of organisms or between a wild type protein and a mutein thereof. *See, e.g.*, GCG Version 6.1. Other programs include FASTA, discussed *supra*.

A preferred algorithm when comparing a sequence of the invention to a database containing a large number of sequences from different organisms is the computer  
 25 program BLAST, especially blastp or tblastn. *See, e.g.*, Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990); Altschul *et al.*, *Nucleic Acids Res.* 25:3389-402 (1997); herein incorporated by reference. Preferred parameters for blastp are:

Expectation value:	10 (default)
Filter:	seg (default)
30 Cost to open a gap:	11 (default)
Cost to extend a gap:	1 (default)

Max. alignments: 100 (default)  
Word size: 11 (default)  
No. of descriptions: 100 (default)  
Penalty Matrix: BLOSUM62

5       The length of polypeptide sequences compared for homology will generally be at least about 16 amino acid residues, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. When searching a database containing sequences from a large number of different organisms, it is preferable to compare amino acid sequences.

10       Database searching using amino acid sequences can be measured by algorithms other than blastp are known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA (*e.g.*, FASTA2 and FASTA3) provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson (1990), *supra*; Pearson (2000),  
15       *supra*. For example, percent sequence identity between amino acid sequences can be determined using FASTA with its default or recommended parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1, herein incorporated by reference.

      An "antibody" refers to an intact immunoglobulin, or to an antigen-binding  
20       portion thereof that competes with the intact antibody for specific binding to a molecular species, *e.g.*, a polypeptide of the instant invention. Antigen-binding portions may be produced by recombinant DNA techniques or by enzymatic or chemical cleavage of intact antibodies. Antigen-binding portions include, *inter alia*, Fab, Fab', F(ab')<sub>2</sub>, Fv, dAb, and complementarity determining region (CDR) fragments, single-chain antibodies  
25       (scFv), chimeric antibodies, diabodies and polypeptides that contain at least a portion of an immunoglobulin that is sufficient to confer specific antigen binding to the polypeptide. An Fab fragment is a monovalent fragment consisting of the VL, VH, CL and CH1 domains; an F(ab')<sub>2</sub> fragment is a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region; an Fd fragment consists of the  
30       VH and CH1 domains; an Fv fragment consists of the VL and VH domains of a single

arm of an antibody; and a dAb fragment consists of a VH domain. *See, e.g., Ward et al., Nature* 341: 544-546 (1989).

By "bind specifically" and "specific binding" is here intended the ability of the antibody to bind to a first molecular species in preference to binding to other molecular species with which the antibody and first molecular species are admixed. An antibody is said specifically to "recognize" a first molecular species when it can bind specifically to that first molecular species.

A single-chain antibody (scFv) is an antibody in which a VL and VH regions are paired to form a monovalent molecules via a synthetic linker that enables them to be made as a single protein chain. *See, e.g., Bird et al., Science* 242: 423-426 (1988); Huston *et al., Proc. Natl. Acad. Sci. USA* 85: 5879-5883 (1988). Diabodies are bivalent, bispecific antibodies in which VH and VL domains are expressed on a single polypeptide chain, but using a linker that is too short to allow for pairing between the two domains on the same chain, thereby forcing the domains to pair with complementary domains of another chain and creating two antigen binding sites. *See e.g., Holliger et al., Proc. Natl. Acad. Sci. USA* 90: 6444-6448 (1993); Poljak *et al., Structure* 2: 1121-1123 (1994). One or more CDRs may be incorporated into a molecule either covalently or noncovalently to make it an immunoadhesin. An immunoadhesin may incorporate the CDR(s) as part of a larger polypeptide chain, may covalently link the CDR(s) to another polypeptide chain, or may incorporate the CDR(s) noncovalently. The CDRs permit the immunoadhesin to specifically bind to a particular antigen of interest. A chimeric antibody is an antibody that contains one or more regions from one antibody and one or more regions from one or more other antibodies.

An antibody may have one or more binding sites. If there is more than one binding site, the binding sites may be identical to one another or may be different. For instance, a naturally-occurring immunoglobulin has two identical binding sites, a single-chain antibody or Fab fragment has one binding site, while a "bispecific" or "bifunctional" antibody has two different binding sites.

An "isolated antibody" is an antibody that (1) is not associated with naturally-associated components, including other naturally-associated antibodies, that accompany it in its native state, (2) is free of other proteins from the same species, (3) is expressed

by a cell from a different species, or (4) does not occur in nature. It is known that purified proteins, including purified antibodies, may be stability with non-naturally-associated components. The non-naturally-associated component may be a protein, such as albumin (e.g., BSA) or a chemical such as polyethylene glycol (PEG).

5           A “neutralizing antibody” or “an inhibitory antibody” is an antibody that inhibits the activity of a polypeptide or blocks the binding of a polypeptide to a ligand that normally binds to it. An “activating antibody” is an antibody that increases the activity of a polypeptide.

          The term “epitope” includes any protein determinant capable of specific binding  
10   to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three dimensional structural characteristics, as well as specific charge characteristics. An antibody is said to specifically bind an antigen when the dissociation constant is  $\leq 1 \mu\text{M}$ , preferably  $\leq 100 \text{ nM}$  and most preferably  $\leq 10 \text{ nM}$ .

15           The term patient includes human and veterinary subjects.

          Throughout this specification and claims, the word “comprise,” or variations such as “comprises” or “comprising,” will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

20           The term “prostate specific” refers to a nucleic acid molecule or polypeptide that is expressed predominantly in the prostate as compared to other tissues in the body. In a preferred embodiment, a “prostate specific” nucleic acid molecule or polypeptide is expressed at a level that is 5-fold higher than any other tissue in the body. In a more preferred embodiment, the “prostate specific” nucleic acid molecule or polypeptide is  
25   expressed at a level that is 10-fold higher than any other tissue in the body, more preferably at least 15-fold, 20-fold, 25-fold, 50-fold or 100-fold higher than any other tissue in the body. Nucleic acid molecule levels may be measured by nucleic acid hybridization, such as Northern blot hybridization, or quantitative PCR. Polypeptide levels may be measured by any method known to accurately quantitate protein levels,  
30   such as Western blot analysis.

Nucleic Acid Molecules, Regulatory Sequences, Vectors,  
Host Cells and Recombinant Methods of Making Polypeptides

*Nucleic Acid Molecules*

5           One aspect of the invention provides isolated nucleic acid molecules that are specific to the prostate or to prostate cells or tissue or that are derived from such nucleic acid molecules. These isolated prostate specific nucleic acids (PSNAs) may be a cDNA, a genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. In a preferred embodiment, the nucleic acid  
10 molecule encodes a polypeptide that is specific to prostate, a prostate-specific polypeptide (PSP). In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO:23-31. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO:1-22.

15           A PSNA may be derived from a human or from another animal. In a preferred embodiment, the PSNA is derived from a human or other mammal. In a more preferred embodiment, the PSNA is derived from a human or other primate. In an even more preferred embodiment, the PSNA is derived from a human.

          In another aspect, the invention provides a nucleic acid molecule that selectively  
20 hybridizes to a nucleic acid molecule encoding a PSNA or a complement thereof. The hybridizing nucleic acid molecule may or may not encode a polypeptide or may not encode a PSP. However, in a preferred embodiment, the hybridizing nucleic acid molecule encodes a PSP. In a more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule that encodes  
25 a polypeptide comprising an amino acid sequence of SEQ ID NO:23-31. In an even more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule comprising the nucleic acid sequence of SEQ ID NO:1-22.

          In a preferred embodiment, the nucleic acid molecule selectively hybridizes to a  
30 nucleic acid molecule encoding a PSP under low stringency conditions. In another preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding a PSP under moderate stringency conditions. In a more preferred

embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding a PSP under high stringency conditions. In an even more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence of SEQ ID NO:23-31. In a yet more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule comprising a nucleic acid sequence selected from SEQ ID NO:1-22. In a preferred embodiment of the invention, the hybridizing nucleic acid molecule may be used to express recombinantly a polypeptide of the invention.

10 In another aspect, the invention provides a nucleic acid molecule that exhibits substantial sequence similarity to a nucleic acid encoding a PSP or a complement of the encoding nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding human PSP. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO:23-31. In a preferred embodiment, the similar nucleic acid molecule is one that has at least 60% sequence identity with a nucleic acid molecule encoding a PSP, such as a polypeptide having an amino acid sequence of SEQ ID NO:23-31, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the similar nucleic acid molecule is one that has at least 90% sequence identity with a nucleic acid molecule encoding a PSP, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a nucleic acid molecule encoding a PSP.

25 In another preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a PSNA or its complement. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule having a nucleic acid sequence of SEQ ID NO:1-22. In a preferred embodiment, the nucleic acid molecule is one that has at least 60% sequence identity with a PSNA, such as one having a nucleic acid sequence of SEQ ID NO:1-22, more preferably at least 70%,

even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the nucleic acid molecule is one that has at least 90% sequence identity with a PSNA, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a PSNA.

A nucleic acid molecule that exhibits substantial sequence similarity may be one that exhibits sequence identity over its entire length to a PSNA or to a nucleic acid molecule encoding a PSP, or may be one that is similar over only a part of its length. In this case, the part is at least 50 nucleotides of the PSNA or the nucleic acid molecule encoding a PSP, preferably at least 100 nucleotides, more preferably at least 150 or 200 nucleotides, even more preferably at least 250 or 300 nucleotides, still more preferably at least 400 or 500 nucleotides.

The substantially similar nucleic acid molecule may be a naturally-occurring one that is derived from another species, especially one derived from another primate, wherein the similar nucleic acid molecule encodes an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO:23-31 or demonstrates significant sequence identity to the nucleotide sequence of SEQ ID NO:1-22. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule from a human, when the PSNA is a member of a gene family. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, hamster, cow, horse and pig; and wild animals, *e.g.*, monkey, fox, lions, tigers, bears, giraffes, zebras, etc. The substantially similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring substantially similar nucleic acid molecule may be isolated directly from humans or other species. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by random mutation of a nucleic acid molecule. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by directed mutation of a PSNA. Further, the substantially



similar nucleic acid molecule may or may not be a PSNA. However, in a preferred embodiment, the substantially similar nucleic acid molecule is a PSNA.

In another embodiment, the invention provides a nucleic acid that is an allelic variant of a PSNA or a nucleic acid encoding a PSP. For instance, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes – more than 1.4 million SNPs have already identified in the human genome, International Human Genome Sequencing Consortium, *Nature* 409: 860-921 (2001) – and the sequence determined from one individual of a species may differ from other allelic forms present within the population. Additionally, small deletions and insertions, rather than single nucleotide polymorphisms, are not uncommon in the general population, and often do not alter the function of the protein. Further, amino acid substitutions occur frequently among natural allelic variants, and often do not substantially change protein function.

In a preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that encodes a PSP. In a more preferred embodiment, the gene is transcribed into an mRNA that encodes a PSP comprising an amino acid sequence of SEQ ID NO:23-31. In another preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that is a PSNA. In a more preferred embodiment, the gene is transcribed into an mRNA that comprises the nucleic acid sequence of SEQ ID NO:1-22. In a preferred embodiment, the allelic variant is a naturally-occurring allelic variant in the species of interest. In a more preferred embodiment, the species of interest is human.

A further object of the invention is to provide a nucleic acid molecule that comprises a part of a nucleic acid sequence of the instant invention. The part may or may not encode a polypeptide, and may or may not encode a polypeptide that is a PSP. However, in a preferred embodiment, the part encodes a PSP. In one aspect, the invention comprises a part of a PSNA. In a second aspect, the invention comprises a part of a nucleic acid molecule that hybridizes or exhibits substantial sequence similarity to a PSNA. In a third aspect, the invention comprises a part of a nucleic acid molecule that is an allelic variant of a PSNA. In a fourth aspect, the invention comprises a part of a nucleic acid molecule that encodes a PSP. A part comprises at least 10 nucleotides, more preferably at least 15, 17, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250,

300, 350, 400 or 500 nucleotides. The maximum size of a nucleic acid part is one nucleotide shorter than the sequence of the nucleic acid molecule encoding the full-length protein.

In another aspect, the invention provides a nucleic acid molecule that encodes a fusion protein, a homologous protein, a polypeptide fragment, a mutein or a polypeptide analog, as described below.

Nucleotide sequences of the instantly-described nucleic acids were determined by sequencing a DNA molecule that had resulted, directly or indirectly, from at least one enzymatic polymerization reaction (*e.g.*, reverse transcription and/or polymerase chain reaction) using an automated sequencer (such as the MegaBACE™ 1000, Molecular Dynamics, Sunnyvale, CA, USA). Further, all amino acid sequences of the polypeptides of the present invention were predicted by translation from the nucleic acid sequences so determined, unless otherwise specified.

In another preferred embodiment, the nucleic acid molecule contains modifications of the native nucleic acid molecule. These modifications include nonnative internucleoside bonds, post-synthetic modifications or altered nucleotide analogues. One having ordinary skill in the art would recognize that the type of modification that can be made will depend upon the intended use of the nucleic acid molecule. For instance, when the nucleic acid molecule is used as a hybridization probe, the range of such modifications will be limited to those that permit sequence-discriminating base pairing of the resulting nucleic acid. When used to direct expression of RNA or protein *in vitro* or *in vivo*, the range of such modifications will be limited to those that permit the nucleic acid to function properly as a polymerization substrate. When the isolated nucleic acid is used as a therapeutic agent, the modifications will be limited to those that do not confer toxicity upon the isolated nucleic acid.

In a preferred embodiment, isolated nucleic acid molecules can include nucleotide analogues that incorporate labels that are directly detectable, such as radiolabels or fluorophores, or nucleotide analogues that incorporate labels that can be visualized in a subsequent reaction, such as biotin or various haptens. In a more preferred embodiment, the labeled nucleic acid molecule may be used as a hybridization probe.

Common radiolabeled analogues include those labeled with  $^{33}\text{P}$ ,  $^{32}\text{P}$ , and  $^{35}\text{S}$ , such as  $^{-32}\text{P}$ -dATP,  $^{-32}\text{P}$ -dCTP,  $^{-32}\text{P}$ -dGTP,  $^{-32}\text{P}$ -dTTP,  $^{-32}\text{P}$ -3'dATP,  $^{-32}\text{P}$ -ATP,  $^{-32}\text{P}$ -CTP,  $^{-32}\text{P}$ -GTP,  $^{-32}\text{P}$ -UTP,  $^{-35}\text{S}$ -dATP,  $^{-35}\text{S}$ -GTP,  $^{-33}\text{P}$ -dATP, and the like.

Commercially available fluorescent nucleotide analogues readily incorporated into the nucleic acids of the present invention include Cy3-dCTP, Cy3-dUTP, Cy5-dCTP, Cy3-dUTP (Amersham Pharmacia Biotech, Piscataway, New Jersey, USA), fluorescein-12-dUTP, tetramethylrhodamine-6-dUTP, Texas Red®-5-dUTP, Cascade Blue®-7-dUTP, BODIPY® FL-14-dUTP, BODIPY® TMR-14-dUTP, BODIPY® TR-14-dUTP, Rhodamine Green™-5-dUTP, Oregon Green® 488-5-dUTP, Texas Red®-12-dUTP, BODIPY® 630/650-14-dUTP, BODIPY® 650/665-14-dUTP, Alexa Fluor® 488-5-dUTP, Alexa Fluor® 532-5-dUTP, Alexa Fluor® 568-5-dUTP, Alexa Fluor® 594-5-dUTP, Alexa Fluor® 546-14-dUTP, fluorescein-12-UTP, tetramethylrhodamine-6-UTP, Texas Red®-5-UTP, Cascade Blue®-7-UTP, BODIPY® FL-14-UTP, BODIPY® TMR-14-UTP, BODIPY® TR-14-UTP, Rhodamine Green™-5-UTP, Alexa Fluor® 488-5-UTP, Alexa Fluor® 546-14-UTP (Molecular Probes, Inc. Eugene, OR, USA). One may also custom synthesize nucleotides having other fluorophores. See Henegariu *et al.*, *Nature Biotechnol.* 18: 345-348 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Haptens that are commonly conjugated to nucleotides for subsequent labeling include biotin (biotin-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA; biotin-21-UTP, biotin-21-dUTP, Clontech Laboratories, Inc., Palo Alto, CA, USA), digoxigenin (DIG-11-dUTP, alkali labile, DIG-11-UTP, Roche Diagnostics Corp., Indianapolis, IN, USA), and dinitrophenyl (dinitrophenyl-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA).

Nucleic acid molecules can be labeled by incorporation of labeled nucleotide analogues into the nucleic acid. Such analogues can be incorporated by enzymatic polymerization, such as by nick translation, random priming, polymerase chain reaction (PCR), terminal transferase tailing, and end-filling of overhangs, for DNA molecules, and *in vitro* transcription driven, *e.g.*, from phage promoters, such as T7, T3, and SP6, for RNA molecules. Commercial kits are readily available for each such labeling approach. Analogues can also be incorporated during automated solid phase chemical synthesis.

Labels can also be incorporated after nucleic acid synthesis, with the 5' phosphate and 3' hydroxyl providing convenient sites for post-synthetic covalent attachment of detectable labels.

Other post-synthetic approaches also permit internal labeling of nucleic acids.

- 5 For example, fluorophores can be attached using a cisplatin reagent that reacts with the N7 of guanine residues (and, to a lesser extent, adenine bases) in DNA, RNA, and PNA to provide a stable coordination complex between the nucleic acid and fluorophore label (Universal Linkage System) (available from Molecular Probes, Inc., Eugene, OR, USA and Amersham Pharmacia Biotech, Piscataway, NJ, USA); *see Alers et al., Genes,*
- 10 *Chromosomes & Cancer* 25: 301- 305 (1999); Jelsma *et al., J. NIH Res.* 5: 82 (1994); Van Belkum *et al., BioTechniques* 16: 148-153 (1994), incorporated herein by reference. As another example, nucleic acids can be labeled using a disulfide-containing linker (FastTag™ Reagent, Vector Laboratories, Inc., Burlingame, CA, USA) that is photo- or thermally coupled to the target nucleic acid using aryl azide chemistry; after reduction, a
- 15 free thiol is available for coupling to a hapten, fluorophore, sugar, affinity ligand, or other marker.

- One or more independent or interacting labels can be incorporated into the nucleic acids of the present invention. For example, both a fluorophore and a moiety that in proximity thereto acts to quench fluorescence can be included to report specific
- 20 hybridization through release of fluorescence quenching or to report exonucleotidic excision. *See, e.g., Tyagi et al., Nature Biotechnol.* 14: 303-308 (1996); Tyagi *et al., Nature Biotechnol.* 16: 49-53 (1998); Sokol *et al., Proc. Natl. Acad. Sci. USA* 95: 11538-11543 (1998); Kostrikis *et al., Science* 279: 1228-1229 (1998); Marras *et al., Genet. Anal.* 14: 151-156 (1999); United States Patent Nos. 5,846,726, 5,925,517,
- 25 5,925,517, 5,723,591 and 5,538,848; Holland *et al., Proc. Natl. Acad. Sci. USA* 88: 7276-7280 (1991); Heid *et al., Genome Res.* 6(10): 986-94 (1996); Kuimelis *et al., Nucleic Acids Symp. Ser.* (37): 255-6 (1997); the disclosures of which are incorporated herein by reference in their entireties.

- Nucleic acid molecules may be modified by altering one or more native
- 30 phosphodiester internucleoside bonds to more nuclease-resistant, internucleoside bonds. *See Hartmann et al. (eds.), Manual of Antisense Methodology: Perspectives in Antisense*

Science, Kluwer Law International (1999); Stein *et al.* (eds.), Applied Antisense Oligonucleotide Technology, Wiley-Liss (1998); Chadwick *et al.* (eds.), Oligonucleotides as Therapeutic Agents - Symposium No. 209, John Wiley & Son Ltd (1997); the disclosures of which are incorporated herein by reference in their entireties.

- 5 Such altered internucleoside bonds are often desired for antisense techniques or for targeted gene correction, Gamper *et al.*, *Nucl. Acids Res.* 28(21): 4332-4339 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Modified oligonucleotide backbones include, without limitation, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Representative United States patents that teach the preparation of the above phosphorus-containing linkages include, but are not limited to, United States Patent Nos. 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050, the disclosures of which are incorporated herein by reference in their entireties. In a preferred embodiment, the modified internucleoside linkages may be used for antisense techniques.

Other modified oligonucleotide backbones do not include a phosphorus atom, but have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and

methylenedihydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH<sub>2</sub> component parts. Representative U.S. patents that teach the preparation of the above backbones include, but are not limited to, United States Patent Nos. 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307; 5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437 and 5,677,439; the disclosures of which are incorporated herein by reference in their entireties.

In other preferred oligonucleotide mimetics, both the sugar and the internucleoside linkage are replaced with novel groups, such as peptide nucleic acids (PNA). In PNA compounds, the phosphodiester backbone of the nucleic acid is replaced with an amide-containing backbone, in particular by repeating N-(2-aminoethyl) glycine units linked by amide bonds. Nucleobases are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone, typically by methylene carbonyl linkages. PNA can be synthesized using a modified peptide synthesis protocol. PNA oligomers can be synthesized by both Fmoc and tBoc methods. Representative U.S. patents that teach the preparation of PNA compounds include, but are not limited to, United States Patent Nos. 5,539,082; 5,714,331; and 5,719,262, each of which is herein incorporated by reference. Automated PNA synthesis is readily achievable on commercial synthesizers (*see, e.g.*, "PNA User's Guide," Rev. 2, February 1998, Perseptive Biosystems Part No. 60138, Applied Biosystems, Inc., Foster City, CA).

PNA molecules are advantageous for a number of reasons. First, because the PNA backbone is uncharged, PNA/DNA and PNA/RNA duplexes have a higher thermal stability than is found in DNA/DNA and DNA/RNA duplexes. The T<sub>m</sub> of a PNA/DNA or PNA/RNA duplex is generally 1°C higher per base pair than the T<sub>m</sub> of the corresponding DNA/DNA or DNA/RNA duplex (in 100 mM NaCl). Second, PNA molecules can also form stable PNA/DNA complexes at low ionic strength, under conditions in which DNA/DNA duplex formation does not occur. Third, PNA also demonstrates greater specificity in binding to complementary DNA because a PNA/DNA mismatch is more destabilizing than DNA/DNA mismatch. A single mismatch in mixed a PNA/DNA 15-mer lowers the T<sub>m</sub> by 8–20°C (15°C on average). In the corresponding

DNA/DNA duplexes, a single mismatch lowers the  $T_m$  by 4–16°C (11°C on average). Because PNA probes can be significantly shorter than DNA probes, their specificity is greater. Fourth, PNA oligomers are resistant to degradation by enzymes, and the lifetime of these compounds is extended both *in vivo* and *in vitro* because nucleases and proteases do not recognize the PNA polyamide backbone with nucleobase sidechains. *See, e.g.,* Ray *et al.*, *FASEB J.* 14(9): 1041-60 (2000); Nielsen *et al.*, *Pharmacol Toxicol.* 86(1): 3-7 (2000); Larsen *et al.*, *Biochim Biophys Acta.* 1489(1): 159-66 (1999); Nielsen, *Curr. Opin. Struct. Biol.* 9(3): 353-7 (1999), and Nielsen, *Curr. Opin. Biotechnol.* 10(1): 71-5 (1999), the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules may be modified compared to their native structure throughout the length of the nucleic acid molecule or can be localized to discrete portions thereof. As an example of the latter, chimeric nucleic acids can be synthesized that have discrete DNA and RNA domains and that can be used for targeted gene repair and modified PCR reactions, as further described in United States Patent Nos. 5,760,012 and 5,731,181, Misra *et al.*, *Biochem.* 37: 1917-1925 (1998); and Finn *et al.*, *Nucl. Acids Res.* 24: 3357-3363 (1996), the disclosures of which are incorporated herein by reference in their entireties.

Unless otherwise specified, nucleic acids of the present invention can include any topological conformation appropriate to the desired use; the term thus explicitly comprehends, among others, single-stranded, double-stranded, triplexed, quadruplexed, partially double-stranded, partially-triplexed, partially-quadruplexed, branched, hairpinned, circular, and padlocked conformations. Padlock conformations and their utilities are further described in Banér *et al.*, *Curr. Opin. Biotechnol.* 12: 11-15 (2001); Escude *et al.*, *Proc. Natl. Acad. Sci. USA* 14: 96(19):10603-7 (1999); Nilsson *et al.*, *Science* 265(5181): 2085-8 (1994), the disclosures of which are incorporated herein by reference in their entireties. Triplex and quadruplex conformations, and their utilities, are reviewed in Praseuth *et al.*, *Biochim. Biophys. Acta.* 1489(1): 181-206 (1999); Fox, *Curr. Med. Chem.* 7(1): 17-37 (2000); Kochetkova *et al.*, *Methods Mol. Biol.* 130: 189-201 (2000); Chan *et al.*, *J. Mol. Med.* 75(4): 267-82 (1997), the disclosures of which are incorporated herein by reference in their entireties.

*Methods for Using Nucleic Acid Molecules as Probes and Primers*

The isolated nucleic acids of the present invention can be used as hybridization probes to detect, characterize, and quantify hybridizing nucleic acids in, and isolate hybridizing nucleic acids from, both genomic and transcript-derived nucleic acid samples. When free in solution, such probes are typically, but not invariably, detectably labeled; bound to a substrate, as in a microarray, such probes are typically, but not invariably unlabeled.

In one embodiment, the isolated nucleic acids of the present invention can be used as probes to detect and characterize gross alterations in the gene of a PSNA, such as deletions, insertions, translocations, and duplications of the PSNA genomic locus through fluorescence *in situ* hybridization (FISH) to chromosome spreads. See, e.g., Andreeff *et al.* (eds.), Introduction to Fluorescence *In Situ* Hybridization: Principles and Clinical Applications, John Wiley & Sons (1999), the disclosure of which is incorporated herein by reference in its entirety. The isolated nucleic acids of the present invention can be used as probes to assess smaller genomic alterations using, e.g., Southern blot detection of restriction fragment length polymorphisms. The isolated nucleic acids of the present invention can be used as probes to isolate genomic clones that include the nucleic acids of the present invention, which thereafter can be restriction mapped and sequenced to identify deletions, insertions, translocations, and substitutions (single nucleotide polymorphisms, SNPs) at the sequence level.

In another embodiment, the isolated nucleic acids of the present invention can be also be used as probes to detect, characterize, and quantify PSNA in, and isolate PSNA from, transcript-derived nucleic acid samples. In one aspect, the isolated nucleic acids of the present invention can be used as hybridization probes to detect, characterize by length, and quantify mRNA by Northern blot of total or poly-A<sup>+</sup>- selected RNA samples. In another aspect, the isolated nucleic acids of the present invention can be used as hybridization probes to detect, characterize by location, and quantify mRNA by *in situ* hybridization to tissue sections. See, e.g., Schwarchzacher *et al.*, In Situ Hybridization, Springer-Verlag New York (2000), the disclosure of which is incorporated herein by reference in its entirety. In another preferred embodiment, the isolated nucleic acids of the present invention can be used as hybridization probes to measure the representation



of clones in a cDNA library or to isolate hybridizing nucleic acid molecules acids from cDNA libraries, permitting sequence level characterization of mRNAs that hybridize to PSNAs, including, without limitations, identification of deletions, insertions, substitutions, truncations, alternatively spliced forms and single nucleotide polymorphisms. In yet another preferred embodiment, the nucleic acid molecules of the instant invention may be used in microarrays.

All of the aforementioned probe techniques are well within the skill in the art, and are described at greater length in standard texts such as Sambrook (2001), *supra*; Ausubel (1999), *supra*; and Walker *et al.* (eds.), The Nucleic Acids Protocols Handbook, Humana Press (2000), the disclosures of which are incorporated herein by reference in their entirety.

Thus, in one embodiment, a nucleic acid molecule of the invention may be used as a probe or primer to identify or amplify a second nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of the invention. In a preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding a PSP. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO:23-31. In another preferred embodiment, the probe or primer is derived from a PSNA. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO:1-22.

In general, a probe or primer is at least 10 nucleotides in length, more preferably at least 12, more preferably at least 14 and even more preferably at least 16 or 17 nucleotides in length. In an even more preferred embodiment, the probe or primer is at least 18 nucleotides in length, even more preferably at least 20 nucleotides and even more preferably at least 22 nucleotides in length. Primers and probes may also be longer in length. For instance, a probe or primer may be 25 nucleotides in length, or may be 30, 40 or 50 nucleotides in length. Methods of performing nucleic acid hybridization using oligonucleotide probes are well-known in the art. *See, e.g.*, Sambrook *et al.*, 1989, *supra*, Chapter 11 and pp. 11.31-11.32 and 11.40-11.44, which describes radiolabeling of short probes, and pp. 11.45-11.53, which describe hybridization conditions for oligonucleotide probes, including specific conditions for probe hybridization (pp. 11.50-11.51).

Methods of performing primer-directed amplification are also well-known in the art. Methods for performing the polymerase chain reaction (PCR) are compiled, *inter alia*, in McPherson, PCR Basics: From Background to Bench, Springer Verlag (2000); Innis *et al.* (eds.), PCR Applications: Protocols for Functional Genomics, Academic Press (1999); Gelfand *et al.* (eds.), PCR Strategies, Academic Press (1998); Newton *et al.*, PCR, Springer-Verlag New York (1997); Burke (ed.), PCR: Essential Techniques, John Wiley & Son Ltd (1996); White (ed.), PCR Cloning Protocols: From Molecular Cloning to Genetic Engineering, Vol. 67, Humana Press (1996); McPherson *et al.* (eds.), PCR 2: A Practical Approach, Oxford University Press, Inc. (1995); the disclosures of which are incorporated herein by reference in their entireties. Methods for performing RT-PCR are collected, *e.g.*, in Siebert *et al.* (eds.), Gene Cloning and Analysis by RT-PCR, Eaton Publishing Company/Bio Techniques Books Division, 1998; Siebert (ed.), PCR Technique: RT-PCR, Eaton Publishing Company/ BioTechniques Books (1995); the disclosure of which is incorporated herein by reference in its entirety.

PCR and hybridization methods may be used to identify and/or isolate allelic variants, homologous nucleic acid molecules and fragments of the nucleic acid molecules of the invention. PCR and hybridization methods may also be used to identify, amplify and/or isolate nucleic acid molecules that encode homologous proteins, analogs, fusion protein or muteins of the invention. The nucleic acid primers of the present invention can be used to prime amplification of nucleic acid molecules of the invention, using transcript-derived or genomic DNA as template.

The nucleic acid primers of the present invention can also be used, for example, to prime single base extension (SBE) for SNP detection (*See, e.g.*, U.S. Pat. No. 6,004,744, the disclosure of which is incorporated herein by reference in its entirety).

Isothermal amplification approaches, such as rolling circle amplification, are also now well-described. *See, e.g.*, Schweitzer *et al.*, *Curr. Opin. Biotechnol.* 12(1): 21-7 (2001); United States Patent Nos. 5,854,033 and 5,714,320; and international patent publications WO 97/19193 and WO 00/15779, the disclosures of which are incorporated herein by reference in their entireties. Rolling circle amplification can be combined with other techniques to facilitate SNP detection. *See, e.g.*, Lizardi *et al.*, *Nature Genet.* 19(3): 225-32 (1998).

Nucleic acid molecules of the present invention may be bound to a substrate either covalently or noncovalently. The substrate can be porous or solid, planar or non-planar, unitary or distributed. The bound nucleic acid molecules may be used as hybridization probes, and may be labeled or unlabeled. In a preferred embodiment, the bound nucleic acid molecules are unlabeled.

In one embodiment, the nucleic acid molecule of the present invention is bound to a porous substrate, *e.g.*, a membrane, typically comprising nitrocellulose, nylon, or positively-charged derivatized nylon. The nucleic acid molecule of the present invention can be used to detect a hybridizing nucleic acid molecule that is present within a labeled nucleic acid sample, *e.g.*, a sample of transcript-derived nucleic acids. In another embodiment, the nucleic acid molecule is bound to a solid substrate, including, without limitation, glass, amorphous silicon, crystalline silicon or plastics. Examples of plastics include, without limitation, polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof. The solid substrate may be any shape, including rectangular, disk-like and spherical. In a preferred embodiment, the solid substrate is a microscope slide or slide-shaped substrate.

The nucleic acid molecule of the present invention can be attached covalently to a surface of the support substrate or applied to a derivatized surface in a chaotropic agent that facilitates denaturation and adherence by presumed noncovalent interactions, or some combination thereof. The nucleic acid molecule of the present invention can be bound to a substrate to which a plurality of other nucleic acids are concurrently bound, hybridization to each of the plurality of bound nucleic acids being separately detectable. At low density, *e.g.* on a porous membrane, these substrate-bound collections are typically denominated macroarrays; at higher density, typically on a solid support, such as glass, these substrate bound collections of plural nucleic acids are colloquially termed microarrays. As used herein, the term microarray includes arrays of all densities. It is, therefore, another aspect of the invention to provide microarrays that include the nucleic acids of the present invention.

*Expression Vectors, Host Cells and Recombinant Methods of Producing Polypeptides*

In another aspect, the present invention provides vectors that comprise one or more of the isolated nucleic acids of the present invention, and host cells in which such  
5 vectors have been introduced.

The vectors can be used, *inter alia*, for propagating the nucleic acids of the present invention in host cells (cloning vectors), for shuttling the nucleic acids of the present invention between host cells derived from disparate organisms (shuttle vectors), for inserting the nucleic acids of the present invention into host cell chromosomes  
10 (insertion vectors), for expressing sense or antisense RNA transcripts of the nucleic acids of the present invention *in vitro* or within a host cell, and for expressing polypeptides encoded by the nucleic acids of the present invention, alone or as fusions to heterologous polypeptides (expression vectors). Vectors of the present invention will often be suitable for several such uses.

15 Vectors are by now well-known in the art, and are described, *inter alia*, in Jones *et al.* (eds.), Vectors: Cloning Applications: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Jones *et al.* (eds.), Vectors: Expression Systems: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Gacesa *et al.*, Vectors: Essential Data, John Wiley & Sons Ltd. (1995); Cid-Arregui  
20 (eds.), Viral Vectors: Basic Science and Gene Therapy, Eaton Publishing Co. (2000); Sambrook (2001), *supra*; Ausubel (1999), *supra*; the disclosures of which are incorporated herein by reference in their entireties. Furthermore, an enormous variety of vectors are available commercially. Use of existing vectors and modifications thereof being well within the skill in the art, only basic features need be described here.

25 Nucleic acid sequences may be expressed by operatively linking them to an expression control sequence in an appropriate expression vector and employing that expression vector to transform an appropriate unicellular host. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Such operative linking of a nucleic sequence of  
30 this invention to an expression control sequence, of course, includes, if not already part

of the nucleic acid sequence, the provision of a translation initiation codon, ATG or GTG, in the correct reading frame upstream of the nucleic acid sequence.

A wide variety of host/expression vector combinations may be employed in expressing the nucleic acid sequences of this invention. Useful expression vectors, for example, may consist of segments of chromosomal, non-chromosomal and synthetic nucleic acid sequences.

In one embodiment, prokaryotic cells may be used with an appropriate vector. Prokaryotic host cells are often used for cloning and expression. In a preferred embodiment, prokaryotic host cells include *E. coli*, *Pseudomonas*, *Bacillus* and *Streptomyces*. In a preferred embodiment, bacterial host cells are used to express the nucleic acid molecules of the instant invention. Useful expression vectors for bacterial hosts include bacterial plasmids, such as those from *E. coli*, *Bacillus* or *Streptomyces*, including pBluescript, pGEX-2T, pUC vectors, col E1, pCR1, pBR322, pMB9 and their derivatives, wider host range plasmids, such as RP4, phage DNAs, e.g., the numerous derivatives of phage lambda, e.g., NM989, GT10 and GT11, and other phages, e.g., M13 and filamentous single stranded phage DNA. Where *E. coli* is used as host, selectable markers are, analogously, chosen for selectivity in gram negative bacteria: e.g., typical markers confer resistance to antibiotics, such as ampicillin, tetracycline, chloramphenicol, kanamycin, streptomycin and zeocin; auxotrophic markers can also be used.

In other embodiments, eukaryotic host cells, such as yeast, insect, mammalian or plant cells, may be used. Yeast cells, typically *S. cerevisiae*, are useful for eukaryotic genetic studies, due to the ease of targeting genetic changes by homologous recombination and the ability to easily complement genetic defects using recombinantly expressed proteins. Yeast cells are useful for identifying interacting protein components, e.g. through use of a two-hybrid system. In a preferred embodiment, yeast cells are useful for protein expression. Vectors of the present invention for use in yeast will typically, but not invariably, contain an origin of replication suitable for use in yeast and a selectable marker that is functional in yeast. Yeast vectors include Yeast Integrating plasmids (e.g., YIp5) and Yeast Replicating plasmids (the YRp and YEp series plasmids), Yeast Centromere plasmids (the YCp series plasmids), Yeast Artificial

Chromosomes (YACs) which are based on yeast linear plasmids, denoted YLp, pGPD-2,  
2 plasmids and derivatives thereof, and improved shuttle vectors such as those  
described in Gietz *et al.*, *Gene*, 74: 527-34 (1988) (YIplac, YEplac and YCplac).  
Selectable markers in yeast vectors include a variety of auxotrophic markers, the most  
5 common of which are (in *Saccharomyces cerevisiae*) URA3, HIS3, LEU2, TRP1 and  
LYS2, which complement specific auxotrophic mutations, such as *ura3-52*, *his3-D1*,  
*leu2-D1*, *trp1-D1* and *lys2-201*.

Insect cells are often chosen for high efficiency protein expression. Where the  
host cells are from *Spodoptera frugiperda* — *e.g.*, Sf9 and Sf21 cell lines, and  
10 expresSFTM cells (Protein Sciences Corp., Meriden, CT, USA) — the vector replicative  
strategy is typically based upon the baculovirus life cycle. Typically, baculovirus  
transfer vectors are used to replace the wild-type AcMNPV polyhedrin gene with a  
heterologous gene of interest. Sequences that flank the polyhedrin gene in the wild-type  
genome are positioned 5' and 3' of the expression cassette on the transfer vectors.  
15 Following co-transfection with AcMNPV DNA, a homologous recombination event  
occurs between these sequences resulting in a recombinant virus carrying the gene of  
interest and the polyhedrin or p10 promoter. Selection can be based upon visual  
screening for lacZ fusion activity.

In another embodiment, the host cells may be mammalian cells, which are  
20 particularly useful for expression of proteins intended as pharmaceutical agents, and for  
screening of potential agonists and antagonists of a protein or a physiological pathway.  
Mammalian vectors intended for autonomous extrachromosomal replication will  
typically include a viral origin, such as the SV40 origin (for replication in cell lines  
expressing the large T-antigen, such as COS1 and COS7 cells), the papillomavirus origin,  
25 or the EBV origin for long term episomal replication (for use, *e.g.*, in 293-EBNA cells,  
which constitutively express the EBV EBNA-1 gene product and adenovirus E1A).  
Vectors intended for integration, and thus replication as part of the mammalian  
chromosome, can, but need not, include an origin of replication functional in mammalian  
cells, such as the SV40 origin. Vectors based upon viruses, such as adenovirus, adeno-  
30 associated virus, vaccinia virus, and various mammalian retroviruses, will typically  
replicate according to the viral replicative strategy. Selectable markers for use in

mammalian cells include resistance to neomycin (G418), blasticidin, hygromycin and to zeocin, and selection based upon the purine salvage pathway using HAT medium.

Expression in mammalian cells can be achieved using a variety of plasmids, including pSV2, pBC12BI, and p91023, as well as lytic virus vectors (*e.g.*, vaccinia virus, adeno virus, and baculovirus), episomal virus vectors (*e.g.*, bovine papillomavirus),  
5 and retroviral vectors (*e.g.*, murine retroviruses). Useful vectors for insect cells include baculoviral vectors and pVL 941.

Plant cells can also be used for expression, with the vector replicon typically derived from a plant virus (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus,  
10 TMV) and selectable markers chosen for suitability in plants.

It is known that codon usage of different host cells may be different. For example, a plant cell and a human cell may exhibit a difference in codon preference for encoding a particular amino acid. As a result, human mRNA may not be efficiently translated in a plant, bacteria or insect host cell. Therefore, another embodiment of this  
15 invention is directed to codon optimization. The codons of the nucleic acid molecules of the invention may be modified to resemble, as much as possible, genes naturally contained within the host cell without altering the amino acid sequence encoded by the nucleic acid molecule.

Any of a wide variety of expression control sequences may be used in these  
20 vectors to express the DNA sequences of this invention. Such useful expression control sequences include the expression control sequences associated with structural genes of the foregoing expression vectors. Expression control sequences that control transcription include, *e.g.*, promoters, enhancers and transcription termination sites. Expression control sequences in eukaryotic cells that control post-transcriptional events include  
25 splice donor and acceptor sites and sequences that modify the half-life of the transcribed RNA, *e.g.*, sequences that direct poly(A) addition or binding sites for RNA-binding proteins. Expression control sequences that control translation include ribosome binding sites, sequences which direct targeted expression of the polypeptide to or within particular cellular compartments, and sequences in the 5' and 3' untranslated regions that  
30 modify the rate or efficiency of translation.

Examples of useful expression control sequences for a prokaryote, *e.g.*, *E. coli*, will include a promoter, often a phage promoter, such as phage lambda pL promoter, the trc promoter, a hybrid derived from the trp and lac promoters, the bacteriophage T7 promoter (in *E. coli* cells engineered to express the T7 polymerase), the TAC or TRC system, the major operator and promoter regions of phage lambda, the control regions of fd coat protein, or the araBAD operon. Prokaryotic expression vectors may further include transcription terminators, such as the aspA terminator, and elements that facilitate translation, such as a consensus ribosome binding site and translation termination codon, Schomer *et al.*, *Proc. Natl. Acad. Sci. USA* 83: 8506-8510 (1986).

Expression control sequences for yeast cells, typically *S. cerevisiae*, will include a yeast promoter, such as the CYC1 promoter, the GAL1 promoter, the GAL10 promoter, ADH1 promoter, the promoters of the yeast  $\alpha$ -mating system, or the GPD promoter, and will typically have elements that facilitate transcription termination, such as the transcription termination signals from the CYC1 or ADH1 gene.

Expression vectors useful for expressing proteins in mammalian cells will include a promoter active in mammalian cells. These promoters include those derived from mammalian viruses, such as the enhancer-promoter sequences from the immediate early gene of the human cytomegalovirus (CMV), the enhancer-promoter sequences from the Rous sarcoma virus long terminal repeat (RSV LTR), the enhancer-promoter from SV40 or the early and late promoters of adenovirus. Other expression control sequences include the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase. Other expression control sequences include those from the gene comprising the PSNA of interest. Often, expression is enhanced by incorporation of polyadenylation sites, such as the late SV40 polyadenylation site and the polyadenylation signal and transcription termination sequences from the bovine growth hormone (BGH) gene, and ribosome binding sites. Furthermore, vectors can include introns, such as intron II of rabbit  $\beta$ -globin gene and the SV40 splice elements.

Preferred nucleic acid vectors also include a selectable or amplifiable marker gene and means for amplifying the copy number of the gene of interest. Such marker genes are well-known in the art. Nucleic acid vectors may also comprise stabilizing sequences (*e.g.*, ori- or ARS-like sequences and telomere-like sequences), or may



alternatively be designed to favor directed or non-directed integration into the host cell genome. In a preferred embodiment, nucleic acid sequences of this invention are inserted in frame into an expression vector that allows high level expression of an RNA which encodes a protein comprising the encoded nucleic acid sequence of interest. Nucleic acid cloning and sequencing methods are well-known to those of skill in the art and are described in an assortment of laboratory manuals, including Sambrook (1989), *supra*, Sambrook (2000), *supra*; and Ausubel (1992), *supra*, Ausubel (1999), *supra*. Product information from manufacturers of biological, chemical and immunological reagents also provide useful information.

- 10           Expression vectors may be either constitutive or inducible. Inducible vectors include either naturally inducible promoters, such as the trc promoter, which is regulated by the lac operon, and the pL promoter, which is regulated by tryptophan, the MMTV-LTR promoter, which is inducible by dexamethasone, or can contain synthetic promoters and/or additional elements that confer inducible control on adjacent promoters.
- 15   Examples of inducible synthetic promoters are the hybrid Plac/ara-1 promoter and the PLtetO-1 promoter. The PLtetO-1 promoter takes advantage of the high expression levels from the PL promoter of phage lambda, but replaces the lambda repressor sites with two copies of operator 2 of the Tn10 tetracycline resistance operon, causing this promoter to be tightly repressed by the Tet repressor protein and induced in response to tetracycline
- 20   (Tc) and Tc derivatives such as anhydrotetracycline. Vectors may also be inducible because they contain hormone response elements, such as the glucocorticoid response element (GRE) and the estrogen response element (ERE), which can confer hormone inducibility where vectors are used for expression in cells having the respective hormone receptors. To reduce background levels of expression, elements responsive to ecdysone,
- 25   an insect hormone, can be used instead, with coexpression of the ecdysone receptor.

- In one aspect of the invention, expression vectors can be designed to fuse the expressed polypeptide to small protein tags that facilitate purification and/or visualization. Tags that facilitate purification include a polyhistidine tag that facilitates purification of the fusion protein by immobilized metal affinity chromatography, for example using NiNTA resin (Qiagen Inc., Valencia, CA, USA) or TALON™ resin
- 30   (cobalt immobilized affinity chromatography medium, Clontech Labs, Palo Alto, CA,

USA). The fusion protein can include a chitin-binding tag and self-excising intein, permitting chitin-based purification with self-removal of the fused tag (IMPACT™ system, New England Biolabs, Inc., Beverly, MA, USA). Alternatively, the fusion protein can include a calmodulin-binding peptide tag, permitting purification by  
5 calmodulin affinity resin (Stratagene, La Jolla, CA, USA), or a specifically excisable fragment of the biotin carboxylase carrier protein, permitting purification of *in vivo* biotinylated protein using an avidin resin and subsequent tag removal (Promega, Madison, WI, USA). As another useful alternative, the proteins of the present invention can be expressed as a fusion to glutathione-S-transferase, the affinity and specificity of  
10 binding to glutathione permitting purification using glutathione affinity resins, such as Glutathione-Superflow Resin (Clontech Laboratories, Palo Alto, CA, USA), with subsequent elution with free glutathione. Other tags include, for example, the Xpress epitope, detectable by anti-Xpress antibody (Invitrogen, Carlsbad, CA, USA), a myc tag, detectable by anti-myc tag antibody, the V5 epitope, detectable by anti-V5 antibody  
15 (Invitrogen, Carlsbad, CA, USA), FLAG® epitope, detectable by anti-FLAG® antibody (Stratagene, La Jolla, CA, USA), and the HA epitope.

For secretion of expressed proteins, vectors can include appropriate sequences that encode secretion signals, such as leader peptides. For example, the pSecTag2  
20 vectors (Invitrogen, Carlsbad, CA, USA) are 5.2 kb mammalian expression vectors that carry the secretion signal from the V-J2-C region of the mouse Ig kappa-chain for efficient secretion of recombinant proteins from a variety of mammalian cell lines.

Expression vectors can also be designed to fuse proteins encoded by the heterologous nucleic acid insert to polypeptides that are larger than purification and/or identification tags. Useful protein fusions include those that permit display of the  
25 encoded protein on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such as those that have a green fluorescent protein (GFP)-like chromophore, fusions to the IgG Fc region, and fusions for use in two hybrid systems.

Vectors for phage display fuse the encoded polypeptide to, *e.g.*, the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous  
30 phage, such as M13. *See* Barbas *et al.*, Phage Display: A Laboratory Manual, Cold Spring Harbor Laboratory Press (2001); Kay *et al.* (eds.), Phage Display of Peptides and

Proteins: A Laboratory Manual, Academic Press, Inc., (1996); Abelson *et al.* (eds.), Combinatorial Chemistry (Methods in Enzymology, Vol. 267) Academic Press (1996). Vectors for yeast display, *e.g.* the pYD1 yeast display vector (Invitrogen, Carlsbad, CA, USA), use the  $\alpha$ -agglutinin yeast adhesion receptor to display recombinant protein on the surface of *S. cerevisiae*. Vectors for mammalian display, *e.g.*, the pDisplay™ vector (Invitrogen, Carlsbad, CA, USA), target recombinant proteins using an N-terminal cell surface targeting signal and a C-terminal transmembrane anchoring domain of platelet derived growth factor receptor.

A wide variety of vectors now exist that fuse proteins encoded by heterologous nucleic acids to the chromophore of the substrate-independent, intrinsically fluorescent green fluorescent protein from *Aequorea victoria* ("GFP") and its variants. The GFP-like chromophore can be selected from GFP-like chromophores found in naturally occurring proteins, such as *A. victoria* GFP (GenBank accession number AAA27721), *Renilla reniformis* GFP, FP583 (GenBank accession no. AF168419) (DsRed), FP593 (AF272711), FP483 (AF168420), FP484 (AF168424), FP595 (AF246709), FP486 (AF168421), FP538 (AF168423), and FP506 (AF168422), and need include only so much of the native protein as is needed to retain the chromophore's intrinsic fluorescence. Methods for determining the minimal domain required for fluorescence are known in the art. *See* Li *et al.*, *J. Biol. Chem.* 272: 28545-28549 (1997). Alternatively, the GFP-like chromophore can be selected from GFP-like chromophores modified from those found in nature. The methods for engineering such modified GFP-like chromophores and testing them for fluorescence activity, both alone and as part of protein fusions, are well-known in the art. *See* Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Palm *et al.*, *Methods Enzymol.* 302: 378-394 (1999), incorporated herein by reference in its entirety. A variety of such modified chromophores are now commercially available and can readily be used in the fusion proteins of the present invention. These include EGFP ("enhanced GFP"), EBFP ("enhanced blue fluorescent protein"), BFP2, EYFP ("enhanced yellow fluorescent protein"), ECFP ("enhanced cyan fluorescent protein") or Citrine. EGFP (*see, e.g.*, Cormack *et al.*, *Gene* 173: 33-38 (1996); United States Patent Nos. 6,090,919 and 5,804,387) is found on a variety of vectors, both plasmid and viral, which are available commercially (Clontech Labs, Palo

Alto, CA, USA); EBFP is optimized for expression in mammalian cells whereas BFP2, which retains the original jellyfish codons, can be expressed in bacteria (*see, e.g., Heim et al., Curr. Biol.* 6: 178-182 (1996) and Cormack *et al., Gene* 173: 33-38 (1996)). Vectors containing these blue-shifted variants are available from Clontech Labs (Palo Alto, CA, USA). Vectors containing EYFP, ECFP (*see, e.g., Heim et al., Curr. Biol.* 6: 178-182 (1996); Miyawaki *et al., Nature* 388: 882-887 (1997)) and Citrine (*see, e.g., Heikal et al., Proc. Natl. Acad. Sci. USA* 97: 11996-12001 (2000)) are also available from Clontech Labs. The GFP-like chromophore can also be drawn from other modified GFPs, including those described in United States Patent Nos. 6,124,128; 6,096,865; 6,090,919; 6,066,476; 6,054,321; 6,027,881; 5,968,750; 5,874,304; 5,804,387; 5,777,079; 5,741,668; and 5,625,048, the disclosures of which are incorporated herein by reference in their entirety. *See also Conn (ed.), Green Fluorescent Protein (Methods in Enzymology, Vol. 302), Academic Press, Inc. (1999).* The GFP-like chromophore of each of these GFP variants can usefully be included in the fusion proteins of the present invention.

Fusions to the IgG Fc region increase serum half life of protein pharmaceutical products through interaction with the FcRn receptor (also denominated the FcRp receptor and the Brambell receptor, FcRb), further described in International Patent Application nos. WO 97/43316, WO 97/34631, WO 96/32478, WO 96/18412.

For long-term, high-yield recombinant production of the proteins, protein fusions, and protein fragments of the present invention, stable expression is preferred. Stable expression is readily achieved by integration into the host cell genome of vectors having selectable markers, followed by selection of these integrants. Vectors such as pUB6/V5-His A, B, and C (Invitrogen, Carlsbad, CA, USA) are designed for high-level stable expression of heterologous proteins in a wide range of mammalian tissue types and cell lines. pUB6/V5-His uses the promoter/enhancer sequence from the human ubiquitin C gene to drive expression of recombinant proteins: expression levels in 293, CHO, and NIH3T3 cells are comparable to levels from the CMV and human EF-1a promoters. The bsd gene permits rapid selection of stably transfected mammalian cells with the potent antibiotic blasticidin.

Replication incompetent retroviral vectors, typically derived from Moloney murine leukemia virus, also are useful for creating stable transfectants having integrated provirus. The highly efficient transduction machinery of retroviruses, coupled with the availability of a variety of packaging cell lines — such as RetroPack™ PT 67, EcoPack2™-293, AmphoPack-293, GP2-293 cell lines (all available from Clontech Laboratories, Palo Alto, CA, USA) — allow a wide host range to be infected with high efficiency; varying the multiplicity of infection readily adjusts the copy number of the integrated provirus.

Of course, not all vectors and expression control sequences will function equally well to express the nucleic acid sequences of this invention. Neither will all hosts function equally well with the same expression system. However, one of skill in the art may make a selection among these vectors, expression control sequences and hosts without undue experimentation and without departing from the scope of this invention. For example, in selecting a vector, the host must be considered because the vector must be replicated in it. The vector's copy number, the ability to control that copy number, the ability to control integration, if any, and the expression of any other proteins encoded by the vector, such as antibiotic or other selection markers, should also be considered. The present invention further includes host cells comprising the vectors of the present invention, either present episomally within the cell or integrated, in whole or in part, into the host cell chromosome. Among other considerations, some of which are described above, a host cell strain may be chosen for its ability to process the expressed protein in the desired fashion. Such post-translational modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation, and it is an aspect of the present invention to provide PSPs with such post-translational modifications.

In selecting an expression control sequence, a variety of factors should also be considered. These include, for example, the relative strength of the sequence, its controllability, and its compatibility with the nucleic acid sequence of this invention, particularly with regard to potential secondary structures. Unicellular hosts should be selected by consideration of their compatibility with the chosen vector, the toxicity of the product coded for by the nucleic acid sequences of this invention, their secretion

characteristics, their ability to fold the polypeptide correctly, their fermentation or culture requirements, and the ease of purification from them of the products coded for by the nucleic acid sequences of this invention.

The recombinant nucleic acid molecules and more particularly, the expression  
5 vectors of this invention may be used to express the polypeptides of this invention as recombinant polypeptides in a heterologous host cell. The polypeptides of this invention may be full-length or less than full-length polypeptide fragments recombinantly expressed from the nucleic acid sequences according to this invention. Such polypeptides include analogs, derivatives and muteins that may or may not have  
10 biological activity.

Vectors of the present invention will also often include elements that permit *in vitro* transcription of RNA from the inserted heterologous nucleic acid. Such vectors typically include a phage promoter, such as that from T7, T3, or SP6, flanking the nucleic acid insert. Often two different such promoters flank the inserted nucleic acid, permitting  
15 separate *in vitro* production of both sense and antisense strands.

Transformation and other methods of introducing nucleic acids into a host cell (e.g., conjugation, protoplast transformation or fusion, transfection, electroporation, liposome delivery, membrane fusion techniques, high velocity DNA-coated pellets, viral infection and protoplast fusion) can be accomplished by a variety of methods which are  
20 well-known in the art (*See*, for instance, Ausubel, *supra*, and Sambrook *et al.*, *supra*). Bacterial, yeast, plant or mammalian cells are transformed or transfected with an expression vector, such as a plasmid, a cosmid, or the like, wherein the expression vector comprises the nucleic acid of interest. Alternatively, the cells may be infected by a viral expression vector comprising the nucleic acid of interest. Depending upon the host cell,  
25 vector, and method of transformation used, transient or stable expression of the polypeptide will be constitutive or inducible. One having ordinary skill in the art will be able to decide whether to express a polypeptide transiently or stably, and whether to express the protein constitutively or inducibly.

A wide variety of unicellular host cells are useful in expressing the DNA  
30 sequences of this invention. These hosts may include well-known eukaryotic and prokaryotic hosts, such as strains of, fungi, yeast, insect cells such as *Spodoptera*

*frugiperda* (SF9), animal cells such as CHO,, as well as plant cells in tissue culture. Representative examples of appropriate host cells include, but are not limited to, bacterial cells, such as *E. coli*, *Caulobacter crescentus*, *Streptomyces* species, and *Salmonella typhimurium*; yeast cells, such as *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pichia pastoris*, *Pichia methanolica*; insect cell lines, such as those from *Spodoptera frugiperda* — e.g., Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA) — *Drosophila* S2 cells, and *Trichoplusia ni* High Five® Cells (Invitrogen, Carlsbad, CA, USA); and mammalian cells. Typical mammalian cells include BHK cells, BSC 1 cells, BSC 40 cells, BMT 10 cells, VERO cells, COS1 cells, COS7 cells, Chinese hamster ovary (CHO) cells, 3T3 cells, NIH 3T3 cells, 293 cells, HEPG2 cells, HeLa cells, L cells, MDCK cells, HEK293 cells, WI38 cells, murine ES cell lines (e.g., from strains 129/SV, C57/BL6, DBA-1, 129/SVJ), K562 cells, Jurkat cells, and BW5147 cells. Other mammalian cell lines are well-known and readily available from the American Type Culture Collection (ATCC) (Manassas, VA, USA) and the National Institute of General medical Sciences (NIGMS) Human Genetic Cell Repository at the Coriell Cell Repositories (Camden, NJ, USA). Cells or cell lines derived from prostate are particularly preferred because they may provide a more native post-translational processing. Particularly preferred are human prostate cells.

Particular details of the transfection, expression and purification of recombinant proteins are well documented and are understood by those of skill in the art. Further details on the various technical aspects of each of the steps used in recombinant production of foreign genes in bacterial cell expression systems can be found in a number of texts and laboratory manuals in the art. See, e.g., Ausubel (1992), *supra*, Ausubel (1999), *supra*, Sambrook (1989), *supra*, and Sambrook (2001), *supra*, herein incorporated by reference.

Methods for introducing the vectors and nucleic acids of the present invention into the host cells are well-known in the art; the choice of technique will depend primarily upon the specific vector to be introduced and the host cell chosen.

Nucleic acid molecules and vectors may be introduced into prokaryotes, such as *E. coli*, in a number of ways. For instance, phage lambda vectors will typically be

packaged using a packaging extract (e.g., Gigapack® packaging extract, Stratagene, La Jolla, CA, USA), and the packaged virus used to infect *E. coli*.

Plasmid vectors will typically be introduced into chemically competent or electrocompetent bacterial cells. *E. coli* cells can be rendered chemically competent by treatment, e.g., with  $\text{CaCl}_2$ , or a solution of  $\text{Mg}^{2+}$ ,  $\text{Mn}^{2+}$ ,  $\text{Ca}^{2+}$ ,  $\text{Rb}^+$  or  $\text{K}^+$ , dimethyl sulfoxide, dithiothreitol, and hexamine cobalt (III), Hanahan, *J. Mol. Biol.* 166(4):557-80 (1983), and vectors introduced by heat shock. A wide variety of chemically competent strains are also available commercially (e.g., Epicurian Coli® XL10-Gold® Ultracompetent Cells (Stratagene, La Jolla, CA, USA); DH5 competent cells (Clontech Laboratories, Palo Alto, CA, USA); and TOP10 Chemically Competent *E. coli* Kit (Invitrogen, Carlsbad, CA, USA)). Bacterial cells can be rendered electrocompetent — that is, competent to take up exogenous DNA by electroporation — by various pre-pulse treatments; vectors are introduced by electroporation followed by subsequent outgrowth in selected media. An extensive series of protocols is provided online in Electroprotocols (BioRad, Richmond, CA, USA) ([http://www.bio-rad.com/LifeScience/pdf/New\\_Gene\\_Pulser.pdf](http://www.bio-rad.com/LifeScience/pdf/New_Gene_Pulser.pdf)).

Vectors can be introduced into yeast cells by spheroplasting, treatment with lithium salts, electroporation, or protoplast fusion. Spheroplasts are prepared by the action of hydrolytic enzymes — a snail-gut extract, usually denoted Glusulase, or Zymolyase, an enzyme from *Arthrobacter luteus* — to remove portions of the cell wall in the presence of osmotic stabilizers, typically 1 M sorbitol. DNA is added to the spheroplasts, and the mixture is co-precipitated with a solution of polyethylene glycol (PEG) and  $\text{Ca}^{2+}$ . Subsequently, the cells are resuspended in a solution of sorbitol, mixed with molten agar and then layered on the surface of a selective plate containing sorbitol.

For lithium-mediated transformation, yeast cells are treated with lithium acetate, which apparently permeabilizes the cell wall, DNA is added and the cells are co-precipitated with PEG. The cells are exposed to a brief heat shock, washed free of PEG and lithium acetate, and subsequently spread on plates containing ordinary selective medium. Increased frequencies of transformation are obtained by using specially-prepared single-stranded carrier DNA and certain organic solvents. Schiestl *et al.*, *Curr. Genet.* 16(5-6): 339-46 (1989).



For electroporation, freshly-grown yeast cultures are typically washed, suspended in an osmotic protectant, such as sorbitol, mixed with DNA, and the cell suspension pulsed in an electroporation device. Subsequently, the cells are spread on the surface of plates containing selective media. Becker *et al.*, *Methods Enzymol.* 194: 182-187 (1991).

- 5 The efficiency of transformation by electroporation can be increased over 100-fold by using PEG, single-stranded carrier DNA and cells that are in late log-phase of growth. Larger constructs, such as YACs, can be introduced by protoplast fusion.

- Mammalian and insect cells can be directly infected by packaged viral vectors, or transfected by chemical or electrical means. For chemical transfection, DNA can be
- 10 coprecipitated with  $\text{CaPO}_4$  or introduced using liposomal and nonliposomal lipid-based agents. Commercial kits are available for  $\text{CaPO}_4$  transfection (CalPhos™ Mammalian Transfection Kit, Clontech Laboratories, Palo Alto, CA, USA), and lipid-mediated transfection can be practiced using commercial reagents, such as LIPOFECTAMINE™ 2000, LIPOFECTAMINE™ Reagent, CELLFECTIN® Reagent, and LIPOFECTIN®
- 15 Reagent (Invitrogen, Carlsbad, CA, USA), DOTAP Liposomal Transfection Reagent, FuGENE 6, X-tremeGENE Q2, DOSPER, (Roche Molecular Biochemicals, Indianapolis, IN USA), Effectene™, PolyFect®, Superfect® (Qiagen, Inc., Valencia, CA, USA). Protocols for electroporating mammalian cells can be found online in Electroprotocols (Bio-Rad, Richmond, CA, USA) ([http://www.bio-](http://www.bio-rad.com/LifeScience/pdf/New_Gene_Pulser.pdf)
- 20 [rad.com/LifeScience/pdf/New\\_Gene\\_Pulser.pdf](http://www.bio-rad.com/LifeScience/pdf/New_Gene_Pulser.pdf)); Norton *et al.* (eds.), Gene Transfer Methods: Introducing DNA into Living Cells and Organisms, BioTechniques Books, Eaton Publishing Co. (2000); incorporated herein by reference in its entirety. Other transfection techniques include transfection by particle bombardment and microinjection. See, e.g., Cheng *et al.*, *Proc. Natl. Acad. Sci. USA* 90(10): 4455-9 (1993); Yang *et al.*,
- 25 *Proc. Natl. Acad. Sci. USA* 87(24): 9568-72 (1990).

Production of the recombinantly produced proteins of the present invention can optionally be followed by purification.

- Purification of recombinantly expressed proteins is now well within the skill in the art. See, e.g., Thorner *et al.* (eds.), Applications of Chimeric Genes and Hybrid
- 30 Proteins, Part A: Gene Expression and Protein Purification (Methods in Enzymology, Vol. 326), Academic Press (2000); Harbin (ed.), Cloning, Gene Expression and Protein

Purification : Experimental Procedures and Process Rationale, Oxford Univ. Press (2001); Marshak *et al.*, Strategies for Protein Purification and Characterization: A Laboratory Course Manual, Cold Spring Harbor Laboratory Press (1996); and Roe (ed.), Protein Purification Applications, Oxford University Press (2001); the disclosures of  
5 which are incorporated herein by reference in their entireties, and thus need not be detailed here.

Briefly, however, if purification tags have been fused through use of an expression vector that appends such tag, purification can be effected, at least in part, by means appropriate to the tag, such as use of immobilized metal affinity chromatography  
10 for polyhistidine tags. Other techniques common in the art include ammonium sulfate fractionation, immunoprecipitation, fast protein liquid chromatography (FPLC), high performance liquid chromatography (HPLC), and preparative gel electrophoresis.

#### Polypeptides

15 Another object of the invention is to provide a polypeptide encoded by a nucleic acid molecule of the instant invention. In a preferred embodiment, the polypeptide is a prostate specific polypeptide (PSP). In an even more preferred embodiment, the polypeptide is derived from a polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:23-31. A polypeptide as defined herein may be  
20 produced recombinantly, as discussed *supra*, may be isolated from a cell that naturally expresses the protein, or may be chemically synthesized following the teachings of the specification and using methods well-known to those having ordinary skill in the art.

In another aspect, the polypeptide may comprise a fragment of a polypeptide, wherein the fragment is as defined herein. In a preferred embodiment, the polypeptide  
25 fragment is a fragment of a PSP. In a more preferred embodiment, the fragment is derived from a polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:23-31. A polypeptide that comprises only a fragment of an entire PSP may or may not be a polypeptide that is also a PSP. For instance, a full-length polypeptide may be prostate-specific, while a fragment thereof may be found in other  
30 tissues as well as in prostate. A polypeptide that is not a PSP, whether it is a fragment, analog, mutein, homologous protein or derivative, is nevertheless useful, especially for

immunizing animals to prepare anti-PSP antibodies. However, in a preferred embodiment, the part or fragment is a PSP. Methods of determining whether a polypeptide is a PSP are described *infra*.

Fragments of at least 6 contiguous amino acids are useful in mapping B cell and T cell epitopes of the reference protein. See, e.g., Geysen *et al.*, *Proc. Natl. Acad. Sci. USA* 81: 3998-4002 (1984) and United States Patent Nos. 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. Because the fragment need not itself be immunogenic, part of an immunodominant epitope, nor even recognized by native antibody, to be useful in such epitope mapping, all fragments of at least 6 amino acids of the proteins of the present invention have utility in such a study.

Fragments of at least 8 contiguous amino acids, often at least 15 contiguous amino acids, are useful as immunogens for raising antibodies that recognize the proteins of the present invention. See, e.g., Lerner, *Nature* 299: 592-596 (1982); Shinnick *et al.*, *Annu. Rev. Microbiol.* 37: 425-46 (1983); Sutcliffe *et al.*, *Science* 219: 660-6 (1983), the disclosures of which are incorporated herein by reference in their entireties. As further described in the above-cited references, virtually all 8-mers, conjugated to a carrier, such as a protein, prove immunogenic — that is, prove capable of eliciting antibody for the conjugated peptide; accordingly, all fragments of at least 8 amino acids of the proteins of the present invention have utility as immunogens.

Fragments of at least 8, 9, 10 or 12 contiguous amino acids are also useful as competitive inhibitors of binding of the entire protein, or a portion thereof, to antibodies (as in epitope mapping), and to natural binding partners, such as subunits in a multimeric complex or to receptors or ligands of the subject protein; this competitive inhibition permits identification and separation of molecules that bind specifically to the protein of interest, United States Patent Nos. 5,539,084 and 5,783,674, incorporated herein by reference in their entireties.

The protein, or protein fragment, of the present invention is thus at least 6 amino acids in length, typically at least 8, 9, 10 or 12 amino acids in length, and often at least 15 amino acids in length. Often, the protein of the present invention, or fragment thereof, is at least 20 amino acids in length, even 25 amino acids, 30 amino acids, 35 amino acids, or 50 amino acids or more in length. Of course, larger fragments having at least 75

amino acids, 100 amino acids, or even 150 amino acids are also useful, and at times preferred.

One having ordinary skill in the art can produce fragments of a polypeptide by truncating the nucleic acid molecule, *e.g.*, a PSNA, encoding the polypeptide and then  
5 expressing it recombinantly. Alternatively, one can produce a fragment by chemically synthesizing a portion of the full-length polypeptide. One may also produce a fragment by enzymatically cleaving either a recombinant polypeptide or an isolated naturally-occurring polypeptide. Methods of producing polypeptide fragments are well-known in the art. *See, e.g.*, Sambrook (1989), *supra*; Sambrook (2001), *supra*; Ausubel (1992),  
10 *supra*; and Ausubel (1999), *supra*. In one embodiment, a polypeptide comprising only a fragment of polypeptide of the invention, preferably a PSP, may be produced by chemical or enzymatic cleavage of a polypeptide. In a preferred embodiment, a polypeptide fragment is produced by expressing a nucleic acid molecule encoding a fragment of the polypeptide, preferably a PSP, in a host cell.

15 *Muteins, Homologous Proteins, Allelic Variants, Analogs and Derivatives*

Another object of the invention is to provide polypeptides that are mutants, fusion proteins, homologous proteins or allelic variants of the polypeptides provided herein.

A mutant protein, or mutein, may have the same or different properties compared to a naturally-occurring polypeptide and comprises at least one amino acid insertion,  
20 duplication, deletion, rearrangement or substitution compared to the amino acid sequence of a native protein. Small deletions and insertions can often be found that do not alter the function of the protein. In one embodiment, the mutein may or may not be prostate-specific. In a preferred embodiment, the mutein is prostate-specific. In a preferred embodiment, the mutein is a polypeptide that comprises at least one amino acid insertion,  
25 duplication, deletion, rearrangement or substitution compared to the amino acid sequence of SEQ ID NO:23-31. In a more preferred embodiment, the mutein is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to a PSP comprising an amino acid sequence of SEQ ID NO:23-31. In a yet more preferred  
30 embodiment, the mutein exhibits at least 85%, more preferably 90%, even more

preferably 95% or 96%, and yet more preferably at least 97%, 98%, 99% or 99.5% sequence identity to a PSP comprising an amino acid sequence of SEQ ID NO:23-31.

A mutein may be produced by isolation from a naturally-occurring mutant cell, tissue or organism. A mutein may be produced by isolation from a cell, tissue or organism that has been experimentally mutagenized. Alternatively, a mutein may be produced by chemical manipulation of a polypeptide, such as by altering the amino acid residue to another amino acid residue using synthetic or semi-synthetic chemical techniques. In a preferred embodiment, a mutein may be produced from a host cell comprising an altered nucleic acid molecule compared to the naturally-occurring nucleic acid molecule. For instance, one may produce a mutein of a polypeptide by introducing one or more mutations into a nucleic acid sequence of the invention and then expressing it recombinantly. These mutations may be targeted, in which particular encoded amino acids are altered, or may be untargeted, in which random encoded amino acids within the polypeptide are altered. Muteins with random amino acid alterations can be screened for a particular biological activity or property, particularly whether the polypeptide is prostate-specific, as described below. Multiple random mutations can be introduced into the gene by methods well-known to the art, *e.g.*, by error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo* mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis and site-specific mutagenesis. Methods of producing muteins with targeted or random amino acid alterations are well-known in the art. *See, e.g.*, Sambrook (1989), *supra*; Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), United States Patent No. 5,223,408, and the references discussed *supra*, each herein incorporated by reference.

The invention also contemplates a polypeptide that is homologous to a polypeptide of the invention. In a preferred embodiment, the polypeptide is homologous to a PSP. In an even more preferred embodiment, the polypeptide is homologous to a PSP selected from the group having an amino acid sequence of SEQ ID NO:23-31. In a preferred embodiment, the homologous polypeptide is one that exhibits significant sequence identity to a PSP. In a more preferred embodiment, the polypeptide is one that exhibits significant sequence identity to an comprising an amino acid sequence of SEQ

ID NO:23-31. In an even more preferred embodiment, the homologous polypeptide is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to a PSP comprising an amino acid sequence of SEQ ID NO:23-31. In a yet  
5 more preferred embodiment, the homologous polypeptide is one that exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97% or 98% sequence identity to a PSP comprising an amino acid sequence of SEQ ID NO:23-31. In another preferred embodiment, the homologous polypeptide is one that exhibits at least 99%, more preferably 99.5%, even more preferably 99.6%,  
10 99.7%, 99.8% or 99.9% sequence identity to a PSP comprising an amino acid sequence of SEQ ID NO:23-31. In a preferred embodiment, the amino acid substitutions are conservative amino acid substitutions as discussed above.

In another embodiment, the homologous polypeptide is one that is encoded by a nucleic acid molecule that selectively hybridizes to a PSNA. In a preferred embodiment,  
15 the homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to a PSNA under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the PSNA is selected from the group consisting of SEQ ID NO:23-31. In another preferred embodiment, the homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to a nucleic acid  
20 molecule that encodes a PSP under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the PSP is selected from the group consisting of SEQ ID NO:23-31.

The homologous polypeptide may be a naturally-occurring one that is derived from another species, especially one derived from another primate, such as chimpanzee,  
25 gorilla, rhesus macaque, baboon or gorilla, wherein the homologous polypeptide comprises an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO:23-31. The homologous polypeptide may also be a naturally-occurring polypeptide from a human, when the PSP is a member of a family of polypeptides. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a  
30 non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, guinea pig, hamster, cow, horse, goat or pig. The

homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring homologous protein may be isolated directly from humans or other species. Alternatively, the nucleic acid molecule encoding the naturally-occurring homologous polypeptide may be isolated and used to express the homologous polypeptide recombinantly. In another embodiment, the homologous polypeptide may be one that is experimentally produced by random mutation of a nucleic acid molecule and subsequent expression of the nucleic acid molecule. In another embodiment, the homologous polypeptide may be one that is experimentally produced by directed mutation of one or more codons to alter the encoded amino acid of a PSP. Further, the homologous protein may or may not encode polypeptide that is a PSP. However, in a preferred embodiment, the homologous polypeptide encodes a polypeptide that is a PSP.

Relatedness of proteins can also be characterized using a second functional test, the ability of a first protein competitively to inhibit the binding of a second protein to an antibody. It is, therefore, another aspect of the present invention to provide isolated proteins not only identical in sequence to those described with particularity herein, but also to provide isolated proteins ("cross-reactive proteins") that competitively inhibit the binding of antibodies to all or to a portion of various of the isolated polypeptides of the present invention. Such competitive inhibition can readily be determined using immunoassays well-known in the art.

As discussed above, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes, and the sequence determined from one individual of a species may differ from other allelic forms present within the population. Thus, in another embodiment, the invention provides a polypeptide encoded by an allelic variant of a nucleic acid molecule encoding a PSP. In a preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that encodes a polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO:23-31. In a yet more preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that has the nucleic acid sequence selected from the group consisting of SEQ ID NO:1-22.

In another embodiment, the invention provides a derivative of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred

embodiment, the polypeptide is a PSP. In a preferred embodiment, the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:23-31, or is a mutein, allelic variant, homologous protein or fragment thereof. In a preferred embodiment, the derivative has been acetylated, carboxylated, phosphorylated, glycosylated or ubiquitinated. In another preferred embodiment, the derivative has been labeled with, *e.g.*, radioactive isotopes such as  $^{125}\text{I}$ ,  $^{32}\text{P}$ ,  $^{35}\text{S}$ , and  $^3\text{H}$ . In another preferred embodiment, the derivative has been labeled with fluorophores, chemiluminescent agents, enzymes, and antiligands that can serve as specific binding pair members for a labeled ligand.

Polypeptide modifications are well-known to those of skill and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, for instance Creighton, Protein Structure and Molecular Properties, 2nd ed., W. H. Freeman and Company (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, in Johnson (ed.), Posttranslational Covalent Modification of Proteins, pgs. 1-12, Academic Press (1983); Seifter *et al.*, *Meth. Enzymol.* 182: 626-646 (1990) and Rattan *et al.*, *Ann. N.Y. Acad. Sci.* 663: 48-62 (1992).

It will be appreciated, as is well-known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well. For



instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

Useful post-synthetic (and post-translational) modifications include conjugation to detectable labels, such as fluorophores. A wide variety of amine-reactive and thiol-reactive fluorophore derivatives have been synthesized that react under non-denaturing  
5 conditions with N-terminal amino groups and epsilon amino groups of lysine residues, on the one hand, and with free thiol groups of cysteine residues, on the other.

Kits are available commercially that permit conjugation of proteins to a variety of amine-reactive or thiol-reactive fluorophores: Molecular Probes, Inc. (Eugene, OR,  
10 USA), *e.g.*, offers kits for conjugating proteins to Alexa Fluor 350, Alexa Fluor 430, Fluorescein-EX, Alexa Fluor 488, Oregon Green 488, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 568, Alexa Fluor 594, and Texas Red-X.

A wide variety of other amine-reactive and thiol-reactive fluorophores are available commercially (Molecular Probes, Inc., Eugene, OR, USA), including Alexa  
15 Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591,  
20 BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA).

The polypeptides of the present invention can also be conjugated to fluorophores,  
25 other proteins, and other macromolecules, using bifunctional linking reagents. Common homobifunctional reagents include, *e.g.*, APG, AEDP, BASED, BMB, BMDB, BMH, BMOE, BM[PEO]3, BM[PEO]4, BS3, BSOCOES, DFDNB, DMA, DMP, DMS, DPDPB, DSG, DSP (Lomant's Reagent), DSS, DST, DTBP, DTME, DTSSP, EGS, HBVS, Sulfo-BSOCOES, Sulfo-DST, Sulfo-EGS (all available from Pierce, Rockford,  
30 IL, USA); common heterobifunctional cross-linkers include ABH, AMAS, ANB-NOS, APDP, ASBA, BMPA, BMPH, BMPS, EDC, EMCA, EMCH, EMCS, KMUA, KMUH,

GMBS, LC-SMCC, LC-SPDP, MBS, M2C2H, MPBH, MSA, NHS-ASA, PDPH, PMPI, SADP, SAED, SAND, SANPAH, SASD, SATP, SBAP, SFAD, SIA, SIAB, SMCC, SMPB, SMPH, SMPT, SPDP, Sulfo-EMCS, Sulfo-GMBS, Sulfo-HSAB, Sulfo-KMUS, Sulfo-LC-SPDP, Sulfo-MBS, Sulfo-NHS-LC-ASA, Sulfo-SADP, Sulfo-SANPAH, 5 Sulfo-SIAB, Sulfo-SMCC, Sulfo-SMPB, Sulfo-LC-SMPT, SVSB, TFCS (all available Pierce, Rockford, IL, USA).

The polypeptides, fragments, and fusion proteins of the present invention can be conjugated, using such cross-linking reagents, to fluorophores that are not amine- or thiol-reactive. Other labels that usefully can be conjugated to the polypeptides, 10 fragments, and fusion proteins of the present invention include radioactive labels, echosonographic contrast reagents, and MRI contrast agents.

The polypeptides, fragments, and fusion proteins of the present invention can also usefully be conjugated using cross-linking agents to carrier proteins, such as KLH, bovine thyroglobulin, and even bovine serum albumin (BSA), to increase 15 immunogenicity for raising anti-PSP antibodies.

The polypeptides, fragments, and fusion proteins of the present invention can also usefully be conjugated to polyethylene glycol (PEG); PEGylation increases the serum half life of proteins administered intravenously for replacement therapy. Delgado *et al.*, *Crit. Rev. Ther. Drug Carrier Syst.* 9(3-4): 249-304 (1992); Scott *et al.*, *Curr. Pharm. Des.* 4(6): 423-38 (1998); DeSantis *et al.*, *Curr. Opin. Biotechnol.* 10(4): 324-30 (1999), 20 incorporated herein by reference in their entireties. PEG monomers can be attached to the protein directly or through a linker, with PEGylation using PEG monomers activated with tresyl chloride (2,2,2-trifluoroethanesulphonyl chloride) permitting direct attachment under mild conditions.

25 In yet another embodiment, the invention provides an analog of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is a PSP. In a more preferred embodiment, the analog is derived from a polypeptide having part or all of the amino acid sequence of SEQ ID NO:23-31. In a preferred embodiment, the analog is one that comprises one or more 30 substitutions of non-natural amino acids or non-native inter-residue bonds compared to the naturally-occurring polypeptide. In general, the non-peptide analog is structurally

similar to a PSP, but one or more peptide linkages is replaced by a linkage selected from the group consisting of  $-\text{CH}_2\text{NH}-$ ,  $-\text{CH}_2\text{S}-$ ,  $-\text{CH}_2\text{-CH}_2-$ ,  $-\text{CH=CH}-$  (cis and trans),  $-\text{COCH}_2-$ ,  $-\text{CH(OH)CH}_2-$  and  $-\text{CH}_2\text{SO}-$ . In another embodiment, the non-peptide analog comprises substitution of one or more amino acids of a PSP with a D-amino acid  
5 of the same type or other non-natural amino acid in order to generate more stable peptides. D-amino acids can readily be incorporated during chemical peptide synthesis: peptides assembled from D-amino acids are more resistant to proteolytic attack; incorporation of D-amino acids can also be used to confer specific three dimensional conformations on the peptide. Other amino acid analogues commonly added during  
10 chemical synthesis include ornithine, norleucine, phosphorylated amino acids (typically phosphoserine, phosphothreonine, phosphotyrosine), L-malonyltyrosine, a non-hydrolyzable analog of phosphotyrosine (*see, e.g., Krole et al., Biochem. Biophys. Res. Com.* 209: 817-821 (1995)), and various halogenated phenylalanine derivatives.

Non-natural amino acids can be incorporated during solid phase chemical  
15 synthesis or by recombinant techniques, although the former is typically more common. Solid phase chemical synthesis of peptides is well established in the art. Procedures are described, inter alia, in Chan *et al.* (eds.), Fmoc Solid Phase Peptide Synthesis: A Practical Approach (Practical Approach Series), Oxford Univ. Press (March 2000); Jones, Amino Acid and Peptide Synthesis (Oxford Chemistry Primers, No 7), Oxford  
20 Univ. Press (1992); and Bodanszky, Principles of Peptide Synthesis (Springer Laboratory), Springer Verlag (1993); the disclosures of which are incorporated herein by reference in their entireties.

Amino acid analogues having detectable labels are also usefully incorporated during synthesis to provide derivatives and analogs. Biotin, for example can be added  
25 using biotinoyl--(9-fluorenylmethoxycarbonyl)-L-lysine (Fmoc biocytin) (Molecular Probes, Eugene, OR, USA). Biotin can also be added enzymatically by incorporation into a fusion protein of a *E. coli* BirA substrate peptide. The Fmoc and tBOC derivatives of dabcyL-L-lysine (Molecular Probes, Inc., Eugene, OR, USA) can be used to incorporate the dabcyL chromophore at selected sites in the peptide sequence during  
30 synthesis. The aminonaphthalene derivative EDANS, the most common fluorophore for

pairing with the dabcyI quencher in fluorescence resonance energy transfer (FRET) systems, can be introduced during automated synthesis of peptides by using EDANS--Fmoc-L-glutamic acid or the corresponding *t*BOC derivative (both from Molecular Probes, Inc., Eugene, OR, USA). Tetramethylrhodamine fluorophores can be  
5 incorporated during automated Fmoc synthesis of peptides using (Fmoc)--TMR-L-lysine (Molecular Probes, Inc. Eugene, OR, USA).

Other useful amino acid analogues that can be incorporated during chemical synthesis include aspartic acid, glutamic acid, lysine, and tyrosine analogues having allyl side-chain protection (Applied Biosystems, Inc., Foster City, CA, USA); the allyl side  
10 chain permits synthesis of cyclic, branched-chain, sulfonated, glycosylated, and phosphorylated peptides.

A large number of other Fmoc-protected non-natural amino acid analogues capable of incorporation during chemical synthesis are available commercially, including, *e.g.*, Fmoc-2-aminobicyclo[2.2.1]heptane-2-carboxylic acid, Fmoc-3-endo-  
15 aminobicyclo[2.2.1]heptane-2-endo-carboxylic acid, Fmoc-3-exo-aminobicyclo[2.2.1]heptane-2-exo-carboxylic acid, Fmoc-3-endo-amino-bicyclo[2.2.1]hept-5-ene-2-endo-carboxylic acid, Fmoc-3-exo-amino-bicyclo[2.2.1]hept-5-ene-2-exo-carboxylic acid, Fmoc-cis-2-amino-1-cyclohexanecarboxylic acid, Fmoc-trans-2-amino-1-cyclohexanecarboxylic acid, Fmoc-1-amino-1-cyclopentanecarboxylic  
20 acid, Fmoc-cis-2-amino-1-cyclopentanecarboxylic acid, Fmoc-1-amino-1-cyclopropanecarboxylic acid, Fmoc-D-2-amino-4-(ethylthio)butyric acid, Fmoc-L-2-amino-4-(ethylthio)butyric acid, Fmoc-L-buthionine, Fmoc-S-methyl-L-Cysteine, Fmoc-2-aminobenzoic acid (anthranillic acid), Fmoc-3-aminobenzoic acid, Fmoc-4-aminobenzoic acid, Fmoc-2-aminobenzophenone-2'-carboxylic acid, Fmoc-N-(4-  
25 aminobenzoyl)- $\beta$ -alanine, Fmoc-2-amino-4,5-dimethoxybenzoic acid, Fmoc-4-aminohippuric acid, Fmoc-2-amino-3-hydroxybenzoic acid, Fmoc-2-amino-5-hydroxybenzoic acid, Fmoc-3-amino-4-hydroxybenzoic acid, Fmoc-4-amino-3-hydroxybenzoic acid, Fmoc-4-amino-2-hydroxybenzoic acid, Fmoc-5-amino-2-hydroxybenzoic acid, Fmoc-2-amino-3-methoxybenzoic acid, Fmoc-4-amino-3-methoxybenzoic acid, Fmoc-2-amino-3-methylbenzoic acid, Fmoc-2-amino-5-methylbenzoic acid, Fmoc-2-amino-6-methylbenzoic acid, Fmoc-3-amino-2-

methylbenzoic acid, Fmoc-3-amino-4-methylbenzoic acid, Fmoc-4-amino-3-  
 methylbenzoic acid, Fmoc-3-amino-2-naphtoic acid, Fmoc-D,L-3-amino-3-  
 phenylpropionic acid, Fmoc-L-Methyldopa, Fmoc-2-amino-4,6-dimethyl-3-  
 pyridinecarboxylic acid, Fmoc-D,L-amino-2-thiophenacetic acid, Fmoc-4-  
 5 (carboxymethyl)piperazine, Fmoc-4-carboxypiperazine, Fmoc-4-  
 (carboxymethyl)homopiperazine, Fmoc-4-phenyl-4-piperidinecarboxylic acid, Fmoc-L-  
 1,2,3,4-tetrahydronorharman-3-carboxylic acid, Fmoc-L-thiazolidine-4-carboxylic acid,  
 all available from The Peptide Laboratory (Richmond, CA, USA).

Non-natural residues can also be added biosynthetically by engineering a  
 10 suppressor tRNA, typically one that recognizes the UAG stop codon, by chemical  
 aminoacylation with the desired unnatural amino acid. Conventional site-directed  
 mutagenesis is used to introduce the chosen stop codon UAG at the site of interest in the  
 protein gene. When the acylated suppressor tRNA and the mutant gene are combined in  
 an *in vitro* transcription/translation system, the unnatural amino acid is incorporated in  
 15 response to the UAG codon to give a protein containing that amino acid at the specified  
 position. Liu *et al.*, *Proc. Natl Acad. Sci. USA* 96(9): 4780-5 (1999); Wang *et al.*,  
*Science* 292(5516): 498-500 (2001).

### *Fusion Proteins*

The present invention further provides fusions of each of the polypeptides and  
 20 fragments of the present invention to heterologous polypeptides. In a preferred  
 embodiment, the polypeptide is a PSP. In a more preferred embodiment, the polypeptide  
 that is fused to the heterologous polypeptide comprises part or all of the amino acid  
 sequence of SEQ ID NO:23-31, or is a mutein, homologous polypeptide, analog or  
 derivative thereof. In an even more preferred embodiment, the nucleic acid molecule  
 25 encoding the fusion protein comprises all or part of the nucleic acid sequence of SEQ ID  
 NO:1-22, or comprises all or part of a nucleic acid sequence that selectively hybridizes or  
 is homologous to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID  
 NO:1-22.

The fusion proteins of the present invention will include at least one fragment of  
 30 the protein of the present invention, which fragment is at least 6, typically at least 8, often

at least 15, and usefully at least 16, 17, 18, 19, or 20 amino acids long. The fragment of the protein of the present to be included in the fusion can usefully be at least 25 amino acids long, at least 50 amino acids long, and can be at least 75, 100, or even 150 amino acids long. Fusions that include the entirety of the proteins of the present invention have particular utility.

The heterologous polypeptide included within the fusion protein of the present invention is at least 6 amino acids in length, often at least 8 amino acids in length, and usefully at least 15, 20, and 25 amino acids in length. Fusions that include larger polypeptides, such as the IgG Fc region, and even entire proteins (such as GFP chromophore-containing proteins) are particular useful.

As described above in the description of vectors and expression vectors of the present invention, which discussion is incorporated here by reference in its entirety, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those designed to facilitate purification and/or visualization of recombinantly-expressed proteins. *See, e.g.*, Ausubel, Chapter 16, (1992), *supra*. Although purification tags can also be incorporated into fusions that are chemically synthesized, chemical synthesis typically provides sufficient purity that further purification by HPLC suffices; however, visualization tags as above described retain their utility even when the protein is produced by chemical synthesis, and when so included render the fusion proteins of the present invention useful as directly detectable markers of the presence of a polypeptide of the invention.

As also discussed above, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those that facilitate secretion of recombinantly expressed proteins — into the periplasmic space or extracellular milieu for prokaryotic hosts, into the culture medium for eukaryotic cells — through incorporation of secretion signals and/or leader sequences. For example, a His<sup>6</sup> tagged protein can be purified on a Ni affinity column and a GST fusion protein can be purified on a glutathione affinity column. Similarly, a fusion protein comprising the Fc domain of IgG can be purified on a Protein A or Protein G column and a fusion protein comprising an epitope tag such as myc can be purified using an immunoaffinity column containing an anti-c-myc antibody. It is preferable that the epitope tag be separated from the protein

encoded by the essential gene by an enzymatic cleavage site that can be cleaved after purification. See also the discussion of nucleic acid molecules encoding fusion proteins that may be expressed on the surface of a cell.

Other useful protein fusions of the present invention include those that permit use of the protein of the present invention as bait in a yeast two-hybrid system. See Bartel *et al.* (eds.), The Yeast Two-Hybrid System, Oxford University Press (1997); Zhu *et al.*, Yeast Hybrid Technologies, Eaton Publishing (2000); Fields *et al.*, *Trends Genet.* 10(8): 286-92 (1994); Mendelsohn *et al.*, *Curr. Opin. Biotechnol.* 5(5): 482-6 (1994); Luban *et al.*, *Curr. Opin. Biotechnol.* 6(1): 59-64 (1995); Allen *et al.*, *Trends Biochem. Sci.* 20(12): 511-6 (1995); Drees, *Curr. Opin. Chem. Biol.* 3(1): 64-70 (1999); Topcu *et al.*, *Pharm. Res.* 17(9): 1049-55 (2000); Fashena *et al.*, *Gene* 250(1-2): 1-14 (2000); ; Colas *et al.*, (1996) Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* 380, 548-550; Norman, T. *et al.*, (1999) Genetic selection of peptide inhibitors of biological pathways. *Science* 285, 591-595, Fabbri *et al.*, (1999) Inhibition of mammalian cell proliferation by genetically selected peptide aptamers that functionally antagonize E2F activity. *Oncogene* 18, 4357-4363; Xu *et al.*, (1997) Cells that register logical relationships among proteins. *Proc Natl Acad Sci U S A.* 94, 12473-12478; Yang, *et al.*, (1995) Protein-peptide interactions analyzed with the yeast two-hybrid system. *Nuc. Acids Res.* 23, 1152-1156; Kolonin *et al.*, (1998) Targeting cyclin-dependent kinases in *Drosophila* with peptide aptamers. *Proc Natl Acad Sci U S A* 95, 14266-14271; Cohen *et al.*, (1998) An artificial cell-cycle inhibitor isolated from a combinatorial library. *Proc Natl Acad Sci U S A* 95, 14272-14277; Uetz, P.; Giot, L.; al, e.; Fields, S.; Rothberg, J. M. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403, 623-627; Ito, *et al.*, (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* 98, 4569-4574, the disclosures of which are incorporated herein by reference in their entireties. Typically, such fusion is to either *E. coli* LexA or yeast GAL4 DNA binding domains. Related bait plasmids are available that express the bait fused to a nuclear localization signal.

Other useful protein fusions include those that permit display of the encoded protein on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such

as green fluorescent protein (GFP), and fusions to the IgG Fc region, as described above, which discussion is incorporated here by reference in its entirety.

The polypeptides and fragments of the present invention can also usefully be fused to protein toxins, such as *Pseudomonas* exotoxin A, diphtheria toxin, shiga toxin A, anthrax toxin lethal factor, ricin, in order to effect ablation of cells that bind or take up the proteins of the present invention.

Fusion partners include, *inter alia*, *myc*, hemagglutinin (HA), GST, immunoglobulins,  $\beta$ -galactosidase, biotin trpE, protein A,  $\beta$ -lactamase,  $\alpha$ -amylase, maltose binding protein, alcohol dehydrogenase, polyhistidine (for example, six histidine at the amino and/or carboxyl terminus of the polypeptide), lacZ, green fluorescent protein (GFP), yeast mating factor, GAL4 transcription activation or DNA binding domain, luciferase, and serum proteins such as ovalbumin, albumin and the constant domain of IgG. See, e.g., Ausubel (1992), *supra* and Ausubel (1999), *supra*. Fusion proteins may also contain sites for specific enzymatic cleavage, such as a site that is recognized by enzymes such as Factor XIII, trypsin, pepsin, or any other enzyme known in the art. Fusion proteins will typically be made by either recombinant nucleic acid methods, as described above, chemically synthesized using techniques well-known in the art (e.g., a Merrifield synthesis), or produced by chemical cross-linking.

Another advantage of fusion proteins is that the epitope tag can be used to bind the fusion protein to a plate or column through an affinity linkage for screening binding proteins or other molecules that bind to the PSP.

As further described below, the isolated polypeptides, muteins, fusion proteins, homologous proteins or allelic variants of the present invention can readily be used as specific immunogens to raise antibodies that specifically recognize PSPs, their allelic variants and homologues. The antibodies, in turn, can be used, *inter alia*, specifically to assay for the polypeptides of the present invention, particularly PSPs, e.g. by ELISA for detection of protein fluid samples, such as serum, by immunohistochemistry or laser scanning cytometry, for detection of protein in tissue samples, or by flow cytometry, for detection of intracellular protein in cell suspensions — for specific antibody-mediated isolation and/or purification of PSPs, as for example by immunoprecipitation, and for use as specific agonists or antagonists of PSPs.



One may determine whether polypeptides, muteins, fusion proteins, homologous proteins or allelic variants are functional by methods known in the art. For instance, residues that are tolerant of change while retaining function can be identified by altering the protein at known residues using methods known in the art, such as alanine scanning mutagenesis, Cunningham *et al.*, *Science* 244(4908): 1081-5 (1989); transposon linker scanning mutagenesis, Chen *et al.*, *Gene* 263(1-2): 39-48 (2001); combinations of homolog- and alanine-scanning mutagenesis, Jin *et al.*, *J. Mol. Biol.* 226(3): 851-65 (1992); combinatorial alanine scanning, Weiss *et al.*, *Proc. Natl. Acad. Sci USA* 97(16): 8950-4 (2000), followed by functional assay. Transposon linker scanning kits are available commercially (New England Biolabs, Beverly, MA, USA, catalog. no. E7-102S; EZ::TN™ In-Frame Linker Insertion Kit, catalogue no. EZI04KN, Epicentre Technologies Corporation, Madison, WI, USA).

Purification of the polypeptides, fragments, homologous polypeptides, muteins, analogs, derivatives and fusion proteins is well-known and within the skill of one having ordinary skill in the art. *See, e.g.*, Scopes, Protein Purification, 2d ed. (1987). Purification of recombinantly expressed polypeptides is described above. Purification of chemically-synthesized peptides can readily be effected, *e.g.*, by HPLC.

Accordingly, it is an aspect of the present invention to provide the isolated proteins of the present invention in pure or substantially pure form in the presence of absence of a stabilizing agent. Stabilizing agents include both proteinaceous or non-proteinaceous material and are well-known in the art. Stabilizing agents, such as albumin and polyethylene glycol (PEG) are known and are commercially available.

Although high levels of purity are preferred when the isolated proteins of the present invention are used as therapeutic agents — such as vaccines, or for replacement therapy — the isolated proteins of the present invention are also useful at lower purity. For example, partially purified proteins of the present invention can be used as immunogens to raise antibodies in laboratory animals.

In preferred embodiments, the purified and substantially purified proteins of the present invention are in compositions that lack detectable ampholytes, acrylamide monomers, bis-acrylamide monomers, and polyacrylamide.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be attached to a substrate. The substrate can be porous or solid, planar or non-planar; the bond can be covalent or noncovalent.

For example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a porous substrate, commonly a membrane, typically comprising nitrocellulose, polyvinylidene fluoride (PVDF), or cationically derivatized, hydrophilic PVDF; so bound, the proteins, fragments, and fusions of the present invention can be used to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention.

As another example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a substantially nonporous substrate, such as plastic, to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention. Such plastics include polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof; when the assay is performed in a standard microtiter dish, the plastic is typically polystyrene.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can also be attached to a substrate suitable for use as a surface enhanced laser desorption ionization source; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biologic interaction there between. The proteins, fragments, and fusions of the present invention can also be attached to a substrate suitable for use in surface plasmon resonance detection; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biological interaction there between.

#### Antibodies

In another aspect, the invention provides antibodies, including fragments and derivatives thereof, that bind specifically to polypeptides encoded by the nucleic acid

molecules of the invention, as well as antibodies that bind to fragments, muteins, derivatives and analogs of the polypeptides. In a preferred embodiment, the antibodies are specific for a polypeptide that is a PSP, or a fragment, mutein, derivative, analog or fusion protein thereof. In a more preferred embodiment, the antibodies are specific for a polypeptide that comprises SEQ ID NO:23-31, or a fragment, mutein, derivative, analog or fusion protein thereof.

The antibodies of the present invention can be specific for linear epitopes, discontinuous epitopes, or conformational epitopes of such proteins or protein fragments, either as present on the protein in its native conformation or, in some cases, as present on the proteins as denatured, as, *e.g.*, by solubilization in SDS. New epitopes may be also due to a difference in post translational modifications (PTMs) in disease versus normal tissue. For example, a particular site on a PSP may be glycosylated in cancerous cells, but not glycosylated in normal cells or vis versa. In addition, alternative splice forms of a PSP may be indicative of cancer. Differential degradation of the C or N-terminus of a PSP may also be a marker or target for anticancer therapy. For example, a PSP may be N-terminal degraded in cancer cells exposing new epitopes to which antibodies may selectively bind for diagnostic or therapeutic uses.

As is well-known in the art, the degree to which an antibody can discriminate as among molecular species in a mixture will depend, in part, upon the conformational relatedness of the species in the mixture; typically, the antibodies of the present invention will discriminate over adventitious binding to non-PSP polypeptides by at least two-fold, more typically by at least 5-fold, typically by more than 10-fold, 25-fold, 50-fold, 75-fold, and often by more than 100-fold, and on occasion by more than 500-fold or 1000-fold. When used to detect the proteins or protein fragments of the present invention, the antibody of the present invention is sufficiently specific when it can be used to determine the presence of the protein of the present invention in samples derived from human prostate.

Typically, the affinity or avidity of an antibody (or antibody multimer, as in the case of an IgM pentamer) of the present invention for a protein or protein fragment of the present invention will be at least about  $1 \times 10^{-6}$  molar (M), typically at least about  $5 \times 10^{-7}$

<sup>7</sup> M,  $1 \times 10^{-7}$  M, with affinities and avidities of at least  $1 \times 10^{-8}$  M,  $5 \times 10^{-9}$  M,  $1 \times 10^{-10}$  M and up to  $1 \times 10^{-13}$  M proving especially useful.

The antibodies of the present invention can be naturally-occurring forms, such as IgG, IgM, IgD, IgE, IgY, and IgA, from any avian, reptilian, or mammalian species.

5 Human antibodies can, but will infrequently, be drawn directly from human donors or human cells. In such case, antibodies to the proteins of the present invention will typically have resulted from fortuitous immunization, such as autoimmune immunization, with the protein or protein fragments of the present invention. Such antibodies will typically, but will not invariably, be polyclonal. In addition, individual  
10 polyclonal antibodies may be isolated and cloned to generate monoclonals.

Human antibodies are more frequently obtained using transgenic animals that express human immunoglobulin genes, which transgenic animals can be affirmatively immunized with the protein immunogen of the present invention. Human Ig-transgenic mice capable of producing human antibodies and methods of producing human  
15 antibodies therefrom upon specific immunization are described, *inter alia*, in United States Patent Nos. 6,162,963; 6,150,584; 6,114,598; 6,075,181; 5,939,598; 5,877,397; 5,874,299; 5,814,318; 5,789,650; 5,770,429; 5,661,016; 5,633,425; 5,625,126; 5,569,825; 5,545,807; 5,545,806, and 5,591,669, the disclosures of which are incorporated herein by reference in their entireties. Such antibodies are typically  
20 monoclonal, and are typically produced using techniques developed for production of murine antibodies.

Human antibodies are particularly useful, and often preferred, when the antibodies of the present invention are to be administered to human beings as *in vivo* diagnostic or therapeutic agents, since recipient immune response to the administered  
25 antibody will often be substantially less than that occasioned by administration of an antibody derived from another species, such as mouse.

IgG, IgM, IgD, IgE, IgY, and IgA antibodies of the present invention are also usefully obtained from other species, including mammals such as rodents — typically mouse, but also rat, guinea pig, and hamster — lagomorphs, typically rabbits, and also  
30 larger mammals, such as sheep, goats, cows, and horses; or other egg laying birds or reptiles such as chickens or alligators. In such cases, as with the transgenic human-

antibody-producing non-human mammals, fortuitous immunization is not required, and the non-human mammal is typically affirmatively immunized, according to standard immunization protocols, with the protein or protein fragment of the present invention.

One form avian antibodies may be generated using techniques described in WO

5 00/29444, published 25 May 2000, the contents of which are hereby incorporated in their entirety.

As discussed above, virtually all fragments of 8 or more contiguous amino acids of the proteins of the present invention can be used effectively as immunogens when conjugated to a carrier, typically a protein such as bovine thyroglobulin, keyhole limpet  
10 hemocyanin, or bovine serum albumin, conveniently using a bifunctional linker such as those described elsewhere above, which discussion is incorporated by reference here.

Immunogenicity can also be conferred by fusion of the polypeptide and fragments of the present invention to other moieties. For example, peptides of the present invention can be produced by solid phase synthesis on a branched polylysine core matrix; these  
15 multiple antigenic peptides (MAPs) provide high purity, increased avidity, accurate chemical definition and improved safety in vaccine development. Tam *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5409-5413 (1988); Posnett *et al.*, *J. Biol. Chem.* 263: 1719-1725 (1988).

Protocols for immunizing non-human mammals or avian species are well-  
20 established in the art. See Harlow *et al.* (eds.), Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory (1998); Coligan *et al.* (eds.), Current Protocols in Immunology, John Wiley & Sons, Inc. (2001); Zola, Monoclonal Antibodies: Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives (Basics: From Background to Bench), Springer Verlag (2000); Gross M, Speck *J.Dtsch. Tierarztl. Wochenschr.* 103: 417-422 (1996), the disclosures of which are incorporated herein by  
25 reference. Immunization protocols often include multiple immunizations, either with or without adjuvants such as Freund's complete adjuvant and Freund's incomplete adjuvant, and may include naked DNA immunization (Moss, *Semin. Immunol.* 2: 317-327 (1990)).

Antibodies from non-human mammals and avian species can be polyclonal or  
30 monoclonal, with polyclonal antibodies having certain advantages in immunohistochemical detection of the proteins of the present invention and monoclonal

antibodies having advantages in identifying and distinguishing particular epitopes of the proteins of the present invention. Antibodies from avian species may have particular advantage in detection of the proteins of the present invention, in human serum or tissues (Viking et al., *Biosens. Bioelectron.* 13: 1257-1262 (1998)).

5       Following immunization, the antibodies of the present invention can be produced using any art-accepted technique. Such techniques are well-known in the art, Coligan, *supra*; Zola, *supra*; Howard *et al.* (eds.), Basic Methods in Antibody Production and Characterization, CRC Press (2000); Harlow, *supra*; Davis (ed.), Monoclonal Antibody Protocols, Vol. 45, Humana Press (1995); Delves (ed.), Antibody Production: Essential  
10 Techniques, John Wiley & Son Ltd (1997); Kenney, Antibody Solution: An Antibody Methods Manual, Chapman & Hall (1997), incorporated herein by reference in their entirety, and thus need not be detailed here.

Briefly, however, such techniques include, *inter alia*, production of monoclonal antibodies by hybridomas and expression of antibodies or fragments or derivatives  
15 thereof from host cells engineered to express immunoglobulin genes or fragments thereof. These two methods of production are not mutually exclusive: genes encoding antibodies specific for the proteins or protein fragments of the present invention can be cloned from hybridomas and thereafter expressed in other host cells. Nor need the two necessarily be performed together: *e.g.*, genes encoding antibodies specific for the  
20 proteins and protein fragments of the present invention can be cloned directly from B cells known to be specific for the desired protein, as further described in United States Patent No. 5,627,052, the disclosure of which is incorporated herein by reference in its entirety, or from antibody-displaying phage.

Recombinant expression in host cells is particularly useful when fragments or  
25 derivatives of the antibodies of the present invention are desired.

Host cells for recombinant antibody production — either whole antibodies, antibody fragments, or antibody derivatives — can be prokaryotic or eukaryotic.

Prokaryotic hosts are particularly useful for producing phage displayed antibodies of the present invention.

30       The technology of phage-displayed antibodies, in which antibody variable region fragments are fused, for example, to the gene III protein (pIII) or gene VIII protein

(pVIII) for display on the surface of filamentous phage, such as M13, is by now well-established. *See, e.g.,* Sidhu, *Curr. Opin. Biotechnol.* 11(6): 610-6 (2000); Griffiths *et al., Curr. Opin. Biotechnol.* 9(1): 102-8 (1998); Hoogenboom *et al., Immunotechnology*, 4(1): 1-20 (1998); Rader *et al., Current Opinion in Biotechnology* 8: 503-508 (1997);  
5 Aujame *et al., Human Antibodies* 8: 155-168 (1997); Hoogenboom, *Trends in Biotechnol.* 15: 62-70 (1997); de Kruif *et al.,* 17: 453-455 (1996); Barbas *et al., Trends in Biotechnol.* 14: 230-234 (1996); Winter *et al., Ann. Rev. Immunol.* 433-455 (1994).  
Techniques and protocols required to generate, propagate, screen (pan), and use the antibody fragments from such libraries have recently been compiled. *See, e.g.,* Barbas  
10 (2001), *supra*; Kay, *supra*; Abelson, *supra*, the disclosures of which are incorporated herein by reference in their entireties.

Typically, phage-displayed antibody fragments are scFv fragments or Fab fragments; when desired, full length antibodies can be produced by cloning the variable regions from the displaying phage into a complete antibody and expressing the full length  
15 antibody in a further prokaryotic or a eukaryotic host cell.

Eukaryotic cells are also useful for expression of the antibodies, antibody fragments, and antibody derivatives of the present invention.

For example, antibody fragments of the present invention can be produced in *Pichia pastoris* and in *Saccharomyces cerevisiae*. *See, e.g.,* Takahashi *et al., Biosci.*  
20 *Biotechnol. Biochem.* 64(10): 2138-44 (2000); Freyre *et al., J. Biotechnol.* 76(2-3):1 57-63 (2000); Fischer *et al., Biotechnol. Appl. Biochem.* 30 (Pt 2): 117-20 (1999); Pennell *et al., Res. Immunol.* 149(6): 599-603 (1998); Eldin *et al., J. Immunol. Methods.* 201(1): 67-75 (1997);, Frenken *et al., Res. Immunol.* 149(6): 589-99 (1998); Shusta *et al., Nature Biotechnol.* 16(8): 773-7 (1998), the disclosures of which are incorporated herein  
25 by reference in their entireties.

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in insect cells. *See, e.g.,* Li *et al., Protein Expr. Purif.*  
21(1): 121-8 (2001); Ailor *et al., Biotechnol. Bioeng.* 58(2-3): 196-203 (1998); Hsu *et al., Biotechnol. Prog.* 13(1): 96-104 (1997); Edelman *et al., Immunology* 91(1): 13-9 (1997);  
30 and Nesbit *et al., J. Immunol. Methods* 151(1-2): 201-8 (1992), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies and fragments and derivatives thereof of the present invention can also be produced in plant cells, particularly maize or tobacco, Giddings *et al.*, *Nature Biotechnol.* 18(11): 1151-5 (2000); Gavilondo *et al.*, *Biotechniques* 29(1): 128-38 (2000); Fischer *et al.*, *J. Biol. Regul. Homeost. Agents* 14(2): 83-92 (2000); Fischer *et al.*,  
5 *Biotechnol. Appl. Biochem.* 30 (Pt 2): 113-6 (1999); Fischer *et al.*, *Biol. Chem.* 380(7-8): 825-39 (1999); Russell, *Curr. Top. Microbiol. Immunol.* 240: 119-38 (1999); and Ma *et al.*, *Plant Physiol.* 109(2): 341-6 (1995), the disclosures of which are incorporated herein by reference in their entirety.

Antibodies, including antibody fragments and derivatives, of the present  
10 invention can also be produced in transgenic, non-human, mammalian milk. *See, e.g.* Pollock *et al.*, *J. Immunol. Methods.* 231: 147-57 (1999); Young *et al.*, *Res. Immunol.* 149: 609-10 (1998); Limonta *et al.*, *Immunotechnology* 1: 107-13 (1995), the disclosures of which are incorporated herein by reference in their entirety.

Mammalian cells useful for recombinant expression of antibodies, antibody  
15 fragments, and antibody derivatives of the present invention include CHO cells, COS cells, 293 cells, and myeloma cells.

Verma *et al.*, *J. Immunol. Methods* 216(1-2):165-81 (1998), herein incorporated by reference, review and compare bacterial, yeast, insect and mammalian expression systems for expression of antibodies.

20 Antibodies of the present invention can also be prepared by cell free translation, as further described in Merk *et al.*, *J. Biochem. (Tokyo)* 125(2): 328-33 (1999) and Ryabova *et al.*, *Nature Biotechnol.* 15(1): 79-84 (1997), and in the milk of transgenic animals, as further described in Pollock *et al.*, *J. Immunol. Methods* 231(1-2): 147-57 (1999), the disclosures of which are incorporated herein by reference in their entirety.

25 The invention further provides antibody fragments that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and  
30 protein fragments encoded by the isolated nucleic acids of the present invention.



Among such useful fragments are Fab, Fab', Fv, F(ab)'<sub>2</sub>, and single chain Fv (scFv) fragments. Other useful fragments are described in Hudson, *Curr. Opin. Biotechnol.* 9(4): 395-402 (1998).

It is also an aspect of the present invention to provide antibody derivatives that  
5 bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or  
10 one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

Among such useful derivatives are chimeric, primatized, and humanized antibodies; such derivatives are less immunogenic in human beings, and thus more suitable for *in vivo* administration, than are unmodified antibodies from non-human mammalian species. Another useful derivative is PEGylation to increase the serum half  
15 life of the antibodies.

Chimeric antibodies typically include heavy and/or light chain variable regions (including both CDR and framework residues) of immunoglobulins of one species, typically mouse, fused to constant regions of another species, typically human. *See, e.g.*, United States Patent No. 5,807,715; Morrison *et al.*, *Proc. Natl. Acad. Sci USA* 81(21):  
20 6851-5 (1984); Sharon *et al.*, *Nature* 309(5966): 364-7 (1984); Takeda *et al.*, *Nature* 314(6010): 452-4 (1985), the disclosures of which are incorporated herein by reference in their entireties. Primatized and humanized antibodies typically include heavy and/or light chain CDRs from a murine antibody grafted into a non-human primate or human antibody V region framework, usually further comprising a human constant region,  
25 Riechmann *et al.*, *Nature* 332(6162): 323-7 (1988); Co *et al.*, *Nature* 351(6326): 501-2 (1991); United States Patent Nos. 6,054,297; 5,821,337; 5,770,196; 5,766,886; 5,821,123; 5,869,619; 6,180,377; 6,013,256; 5,693,761; and 6,180,370, the disclosures of which are incorporated herein by reference in their entireties.

Other useful antibody derivatives of the invention include heteromeric antibody  
30 complexes and antibody fusions, such as diabodies (bispecific antibodies), single-chain diabodies, and intrabodies.

It is contemplated that the nucleic acids encoding the antibodies of the present invention can be operably joined to other nucleic acids forming a recombinant vector for cloning or for expression of the antibodies of the invention. The present invention includes any recombinant vector containing the coding sequences, or part thereof, whether for eukaryotic transduction, transfection or gene therapy. Such vectors may be prepared using conventional molecular biology techniques, known to those with skill in the art, and would comprise DNA encoding sequences for the immunoglobulin V-regions including framework and CDRs or parts thereof, and a suitable promoter either with or without a signal sequence for intracellular transport. Such vectors may be transduced or transfected into eukaryotic cells or used for gene therapy (Marasco et al., *Proc. Natl. Acad. Sci. (USA)* 90: 7889-7893 (1993); Duan et al., *Proc. Natl. Acad. Sci. (USA)* 91: 5075-5079 (1994), by conventional techniques, known to those with skill in the art.

The antibodies of the present invention, including fragments and derivatives thereof, can usefully be labeled. It is, therefore, another aspect of the present invention to provide labeled antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

The choice of label depends, in part, upon the desired use.

For example, when the antibodies of the present invention are used for immunohistochemical staining of tissue samples, the label can usefully be an enzyme that catalyzes production and local deposition of a detectable product.

Enzymes typically conjugated to antibodies to permit their immunohistochemical visualization are well-known, and include alkaline phosphatase,  $\beta$ -galactosidase, glucose oxidase, horseradish peroxidase (HRP), and urease. Typical substrates for production and deposition of visually detectable products include o-nitrophenyl-beta-D-galactopyranoside (ONPG); o-phenylenediamine dihydrochloride (OPD); p-nitrophenyl phosphate (PNPP); p-nitrophenyl-beta-D-galactopyranoside (PNPG); 3',3'-diaminobenzidine (DAB); 3-amino-9-ethylcarbazole (AEC);

4-chloro-1-naphthol (CN); 5-bromo-4-chloro-3-indolyl-phosphate (BCIP); ABTS®; BlueGal; iodonitrotetrazolium (INT); nitroblue tetrazolium chloride (NBT); phenazine methosulfate (PMS); phenolphthalein monophosphate (PMP); tetramethyl benzidine (TMB); tetranitroblue tetrazolium (TNBT); X-Gal; X-Gluc; and X-Glucoside.

- 5 Other substrates can be used to produce products for local deposition that are luminescent. For example, in the presence of hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>), horseradish peroxidase (HRP) can catalyze the oxidation of cyclic diacylhydrazides, such as luminol. Immediately following the oxidation, the luminol is in an excited state (intermediate reaction product), which decays to the ground state by emitting light. Strong
- 10 enhancement of the light emission is produced by enhancers, such as phenolic compounds. Advantages include high sensitivity, high resolution, and rapid detection without radioactivity and requiring only small amounts of antibody. *See, e.g., Thorpe et al., Methods Enzymol.* 133: 331-53 (1986); Kricka *et al., J. Immunoassay* 17(1): 67-83 (1996); and Lundqvist *et al., J. Biolumin. Chemilumin.* 10(6): 353-9 (1995), the
- 15 disclosures of which are incorporated herein by reference in their entireties. Kits for such enhanced chemiluminescent detection (ECL) are available commercially.

The antibodies can also be labeled using colloidal gold.

- As another example, when the antibodies of the present invention are used, *e.g.,* for flow cytometric detection, for scanning laser cytometric detection, or for fluorescent
- 20 immunoassay, they can usefully be labeled with fluorophores.

There are a wide variety of fluorophore labels that can usefully be attached to the antibodies of the present invention.

- For flow cytometric applications, both for extracellular detection and for intracellular detection, common useful fluorophores can be fluorescein isothiocyanate
- 25 (FITC), allophycocyanin (APC), R-phycoerythrin (PE), peridinin chlorophyll protein (PerCP), Texas Red, Cy3, Cy5, fluorescence resonance energy tandem fluorophores such as PerCP-Cy5.5, PE-Cy5, PE-Cy5.5, PE-Cy7, PE-Texas Red, and APC-Cy7.

- Other fluorophores include, *inter alia*, Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa
- 30 Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY

R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA), and Cy2, Cy3, Cy3.5, Cy5, Cy5.5, Cy7, all of which are also useful for fluorescently labeling the antibodies of the present invention.

For secondary detection using labeled avidin, streptavidin, captavidin or neutravidin, the antibodies of the present invention can usefully be labeled with biotin.

When the antibodies of the present invention are used, *e.g.*, for western blotting applications, they can usefully be labeled with radioisotopes, such as  $^{33}\text{P}$ ,  $^{32}\text{P}$ ,  $^{35}\text{S}$ ,  $^3\text{H}$ , and  $^{125}\text{I}$ .

As another example, when the antibodies of the present invention are used for radioimmunotherapy, the label can usefully be  $^{228}\text{Th}$ ,  $^{227}\text{Ac}$ ,  $^{225}\text{Ac}$ ,  $^{223}\text{Ra}$ ,  $^{213}\text{Bi}$ ,  $^{212}\text{Pb}$ ,  $^{212}\text{Bi}$ ,  $^{211}\text{At}$ ,  $^{203}\text{Pb}$ ,  $^{194}\text{Os}$ ,  $^{188}\text{Re}$ ,  $^{186}\text{Re}$ ,  $^{153}\text{Sm}$ ,  $^{149}\text{Tb}$ ,  $^{131}\text{I}$ ,  $^{125}\text{I}$ ,  $^{111}\text{In}$ ,  $^{105}\text{Rh}$ ,  $^{99\text{m}}\text{Tc}$ ,  $^{97}\text{Ru}$ ,  $^{90}\text{Y}$ ,  $^{90}\text{Sr}$ ,  $^{88}\text{Y}$ ,  $^{72}\text{Se}$ ,  $^{67}\text{Cu}$ , or  $^{47}\text{Sc}$ .

As another example, when the antibodies of the present invention are to be used for *in vivo* diagnostic use, they can be rendered detectable by conjugation to MRI contrast agents, such as gadolinium diethylenetriaminepentaacetic acid (DTPA), Lauffer *et al.*, *Radiology* 207(2): 529-38 (1998), or by radioisotopic labeling.

As would be understood, use of the labels described above is not restricted to the application as for which they were mentioned.

The antibodies of the present invention, including fragments and derivatives thereof, can also be conjugated to toxins, in order to target the toxin's ablative action to cells that display and/or express the proteins of the present invention. Commonly, the antibody in such immunotoxins is conjugated to *Pseudomonas* exotoxin A, diphtheria toxin, shiga toxin A, anthrax toxin lethal factor, or ricin. See Hall (ed.), Immunotoxin Methods and Protocols (Methods in Molecular Biology, vol. 166), Humana Press (2000); and Frankel *et al.* (eds.), Clinical Applications of Immunotoxins, Springer-Verlag (1998), the disclosures of which are incorporated herein by reference in their entireties.

The antibodies of the present invention can usefully be attached to a substrate, and it is, therefore, another aspect of the invention to provide antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, attached to a substrate.

Substrates can be porous or nonporous, planar or nonplanar.

For example, the antibodies of the present invention can usefully be conjugated to filtration media, such as NHS-activated Sepharose or CNBr-activated Sepharose for purposes of immunoaffinity chromatography.

For example, the antibodies of the present invention can usefully be attached to paramagnetic microspheres, typically by biotin-streptavidin interaction, which microsphere can then be used for isolation of cells that express or display the proteins of the present invention. As another example, the antibodies of the present invention can usefully be attached to the surface of a microtiter plate for ELISA.

As noted above, the antibodies of the present invention can be produced in prokaryotic and eukaryotic cells. It is, therefore, another aspect of the present invention to provide cells that express the antibodies of the present invention, including hybridoma cells, B cells, plasma cells, and host cells recombinantly modified to express the antibodies of the present invention.

In yet a further aspect, the present invention provides aptamers evolved to bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

In sum, one of skill in the art, provided with the teachings of this invention, has available a variety of methods which may be used to alter the biological properties of the

antibodies of this invention including methods which would increase or decrease the stability or half-life, immunogenicity, toxicity, affinity or yield of a given antibody molecule, or to alter it in any other way that may render it more suitable for a particular application.

5        Transgenic Animals and Cells

In another aspect, the invention provides transgenic cells and non-human organisms comprising nucleic acid molecules of the invention. In a preferred embodiment, the transgenic cells and non-human organisms comprise a nucleic acid molecule encoding a PSP. In a preferred embodiment, the PSP comprises an amino acid  
10        sequence selected from SEQ ID NO:23-31, or a fragment, mutein, homologous protein or allelic variant thereof. In another preferred embodiment, the transgenic cells and non-human organism comprise a PSNA of the invention, preferably a PSNA comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:1-22, or a part,  
15        substantially similar nucleic acid molecule, allelic variant or hybridizing nucleic acid molecule thereof.

In another embodiment, the transgenic cells and non-human organisms have a targeted disruption or replacement of the endogenous orthologue of the huma PSG. The transgenic cells can be embryonic stem cells or somatic cells. The transgenic non-human  
20        organisms can be chimeric, nonchimeric heterozygotes, and nonchimeric homozygotes. Methods of producing transgenic animals are well-known in the art. *See, e.g.,* Hogan *et al.*, Manipulating the Mouse Embryo: A Laboratory Manual, 2d ed., Cold Spring Harbor Press (1999); Jackson *et al.*, Mouse Genetics and Transgenics: A Practical Approach, Oxford University Press (2000); and Pinkert, Transgenic Animal Technology: A  
25        Laboratory Handbook, Academic Press (1999).

Any technique known in the art may be used to introduce nucleic acid molecule of the invention into an animal to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection. (*see, e.g.,* Paterson  
30        *et al.*, *Appl. Microbiol. Biotechnol.* 40: 691-698 (1994); Carver *et al.*, *Biotechnology* 11: 1263-1270 (1993); Wright *et al.*, *Biotechnology* 9: 830-834 (1991); and United States Patent No. 4,873,191 (1989 retrovirus-mediated gene transfer into germ lines, blastocysts

or embryos (*see, e.g., Van der Putten et al., Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985)); gene targeting in embryonic stem cells (*see, e.g., Thompson et al., Cell* 56: 313-321 (1989)); electroporation of cells or embryos (*see, e.g., Lo, 1983, Mol. Cell. Biol.* 3: 1803-1814 (1983)); introduction using a gene gun (*see, e.g., Ulmer et al., Science* 259: 1745-49 (1993)); introducing nucleic acid constructs into embryonic pluripotent stem cells and transferring the stem cells back into the blastocyst; and sperm-mediated gene transfer (*see, e.g., Lavitrano et al., Cell* 57: 717-723 (1989)).

Other techniques include, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (*see, e.g., Campell et al., Nature* 380: 64-66 (1996); Wilmut et al., *Nature* 385: 810-813 (1997)). The present invention provides for transgenic animals that carry the transgene (*i.e., a* nucleic acid molecule of the invention) in all their cells, as well as animals which carry the transgene in some, but not all their cells, *i. e.,* mosaic animals or chimeric animals.

The transgene may be integrated as a single transgene or as multiple copies, such as in concatamers, *e. g.,* head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, *e.g.,* the teaching of Lasko et al. et al., *Proc. Natl. Acad. Sci. USA* 89: 6232-6236 (1992). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and reverse transcriptase-PCR (RT-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Methods for creating a transgenic animal with a disruption of a targeted gene are also well-known in the art. In general, a vector is designed to comprise some nucleotide sequences homologous to the endogenous targeted gene. The vector is introduced into a cell so that it may integrate, via homologous recombination with chromosomal sequences, into the endogenous gene, thereby disrupting the function of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type. *See, e.g., Gu et al., Science* 265: 103-106 (1994). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. *See, e.g., Smithies et al., Nature* 317: 230-234 (1985); Thomas *et al., Cell* 51: 503-512 (1987); Thompson *et al., Cell* 5: 313-321 (1989).

In one embodiment, a mutant, non-functional nucleic acid molecule of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous nucleic acid sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable



marker, to transfect cells that express polypeptides of the invention in vivo. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene. *See, e.g., Thomas, supra* and *Thompson, supra*. However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site in vivo using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (*e.g.,* knockouts) are administered to a patient in vivo. Such cells may be obtained from an animal or patient or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (*e.g.,* lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered in vitro using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, *e.g.,* by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, *e.g.,* in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, *e.g.,* genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or

vascular graft. *See, e.g.*, United States Patent Nos. 5,399,349; 5,460,959, each of which is incorporated by reference herein in its entirety.

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well-known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

#### Computer Readable Means

A further aspect of the invention is a computer readable means for storing the nucleic acid and amino acid sequences of the instant invention. In a preferred embodiment, the invention provides a computer readable means for storing SEQ ID NO:23-31 and SEQ ID NO:1-22 as described herein, as the complete set of sequences or in any combination. The records of the computer readable means can be accessed for reading and display and for interface with a computer system for the application of programs allowing for the location of data upon a query for data meeting certain criteria, the comparison of sequences, the alignment or ordering of sequences meeting a set of criteria, and the like.

The nucleic acid and amino acid sequences of the invention are particularly useful as components in databases useful for search analyses as well as in sequence analysis algorithms. As used herein, the terms "nucleic acid sequences of the invention" and "amino acid sequences of the invention" mean any detectable chemical or physical characteristic of a polynucleotide or polypeptide of the invention that is or may be reduced to or stored in a computer readable form. These include, without limitation,

chromatographic scan data or peak data, photographic data or scan data therefrom, and mass spectrographic data.

This invention provides computer readable media having stored thereon sequences of the invention. A computer readable medium may comprise one or more of the following: a nucleic acid sequence comprising a sequence of a nucleic acid sequence of the invention; an amino acid sequence comprising an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of one or more nucleic acid sequences of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of a nucleic acid sequence of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention. The computer readable medium can be any composition of matter used to store information or data, including, for example, commercially available floppy disks, tapes, hard drives, compact disks, and video disks.

Also provided by the invention are methods for the analysis of character sequences, particularly genetic sequences. Preferred methods of sequence analysis include, for example, methods of sequence homology analysis, such as identity and similarity analysis, RNA structure analysis, sequence assembly, cladistic analysis, sequence motif analysis, open reading frame determination, nucleic acid base calling, and sequencing chromatogram peak analysis.

A computer-based method is provided for performing nucleic acid sequence identity or similarity identification. This method comprises the steps of providing a nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a

computer readable medium; and comparing said nucleic acid sequence to at least one nucleic acid or amino acid sequence to identify sequence identity or similarity.

A computer-based method is also provided for performing amino acid homology identification, said method comprising the steps of: providing an amino acid sequence  
5 comprising the sequence of an amino acid of the invention in a computer readable medium; and comparing said an amino acid sequence to at least one nucleic acid or an amino acid sequence to identify homology.

A computer-based method is still further provided for assembly of overlapping nucleic acid sequences into a single nucleic acid sequence, said method comprising the  
10 steps of: providing a first nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and screening for at least one overlapping region between said first nucleic acid sequence and a second nucleic acid sequence.

#### Diagnostic Methods for Prostate Cancer

15 The present invention also relates to quantitative and qualitative diagnostic assays and methods for detecting, diagnosing, monitoring, staging and predicting cancers by comparing expression of a PSNA or a PSP in a human patient that has or may have prostate cancer, or who is at risk of developing prostate cancer, with the expression of a  
20 PSNA or a PSP in a normal human control. For purposes of the present invention, "expression of a PSNA" or "PSNA expression" means the quantity of PSG mRNA that can be measured by any method known in the art or the level of transcription that can be measured by any method known in the art in a cell, tissue, organ or whole patient. Similarly, the term "expression of a PSP" or "PSP expression" means the amount of PSP  
25 that can be measured by any method known in the art or the level of translation of a PSG PSNA that can be measured by any method known in the art.

The present invention provides methods for diagnosing prostate cancer in a patient, in particular squamous cell carcinoma, by analyzing for changes in levels of PSNA or PSP in cells, tissues, organs or bodily fluids compared with levels of PSNA or  
30 PSP in cells, tissues, organs or bodily fluids of preferably the same type from a normal human control, wherein an increase, or decrease in certain cases, in levels of a PSNA or

PSP in the patient versus the normal human control is associated with the presence of prostate cancer or with a predilection to the disease. In another preferred embodiment, the present invention provides methods for diagnosing prostate cancer in a patient by analyzing changes in the structure of the mRNA of a PSG compared to the mRNA from a normal control. These changes include, without limitation, aberrant splicing, alterations in polyadenylation and/or alterations in 5' nucleotide capping. In yet another preferred embodiment, the present invention provides methods for diagnosing prostate cancer in a patient by analyzing changes in a PSP compared to a PSP from a normal control. These changes include, *e.g.*, alterations in glycosylation and/or phosphorylation of the PSP or subcellular PSP localization.

In a preferred embodiment, the expression of a PSNA is measured by determining the amount of an mRNA that encodes an amino acid sequence selected from SEQ ID NO:23-31, a homolog, an allelic variant, or a fragment thereof. In a more preferred embodiment, the PSNA expression that is measured is the level of expression of a PSNA mRNA selected from SEQ ID NO:1-22, or a hybridizing nucleic acid, homologous nucleic acid or allelic variant thereof, or a part of any of these nucleic acids. PSNA expression may be measured by any method known in the art, such as those described *supra*, including measuring mRNA expression by Northern blot, quantitative or qualitative reverse transcriptase PCR (RT-PCR), microarray, dot or slot blots or *in situ* hybridization. *See, e.g.*, Ausubel (1992), *supra*; Ausubel (1999), *supra*; Sambrook (1989), *supra*; and Sambrook (2001), *supra*. PSNA transcription may be measured by any method known in the art including using a reporter gene hooked up to the promoter of a PSG of interest or doing nuclear run-off assays. Alterations in mRNA structure, *e.g.*, aberrant splicing variants, may be determined by any method known in the art, including, RT-PCR followed by sequencing or restriction analysis. As necessary, PSNA expression may be compared to a known control, such as normal prostate nucleic acid, to detect a change in expression.

In another preferred embodiment, the expression of a PSP is measured by determining the level of a PSP having an amino acid sequence selected from the group consisting of SEQ ID NO:23-31, a homolog, an allelic variant, or a fragment thereof. Such levels are preferably determined in at least one of cells, tissues, organs and/or

bodily fluids, including determination of normal and abnormal levels. Thus, for instance, a diagnostic assay in accordance with the invention for diagnosing over- or underexpression of PSNA or PSP compared to normal control bodily fluids, cells, or tissue samples may be used to diagnose the presence of prostate cancer. The expression  
5 level of a PSP may be determined by any method known in the art, such as those described *supra*. In a preferred embodiment, the PSP expression level may be determined by radioimmunoassays, competitive-binding assays, ELISA, Western blot, FACS, immunohistochemistry, immunoprecipitation, proteomic approaches: two-dimensional gel electrophoresis (2D electrophoresis) and non-gel-based approaches  
10 such as mass spectrometry or protein interaction profiling. *See, e.g.,* Harlow (1999), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. Alterations in the PSP structure may be determined by any method known in the art, including, *e.g.,* using antibodies that specifically recognize phosphoserine, phosphothreonine or phosphotyrosine residues, two-dimensional polyacrylamide gel electrophoresis (2D  
15 PAGE) and/or chemical analysis of amino acid residues of the protein. *Id.*

In a preferred embodiment, a radioimmunoassay (RIA) or an ELISA is used. An antibody specific to a PSP is prepared if one is not already available. In a preferred embodiment, the antibody is a monoclonal antibody. The anti-PSP antibody is bound to a solid support and any free protein binding sites on the solid support are blocked with a  
20 protein such as bovine serum albumin. A sample of interest is incubated with the antibody on the solid support under conditions in which the PSP will bind to the anti-PSP antibody. The sample is removed, the solid support is washed to remove unbound material, and an anti-PSP antibody that is linked to a detectable reagent (a radioactive substance for RIA and an enzyme for ELISA) is added to the solid support and incubated  
25 under conditions in which binding of the PSP to the labeled antibody will occur. After binding, the unbound labeled antibody is removed by washing. For an ELISA, one or more substrates are added to produce a colored reaction product that is based upon the amount of a PSP in the sample. For an RIA, the solid support is counted for radioactive decay signals by any method known in the art. Quantitative results for both RIA and  
30 ELISA typically are obtained by reference to a standard curve.

Other methods to measure PSP levels are known in the art. For instance, a competition assay may be employed wherein an anti-PSP antibody is attached to a solid support and an allocated amount of a labeled PSP and a sample of interest are incubated with the solid support. The amount of labeled PSP detected which is attached to the solid support can be correlated to the quantity of a PSP in the sample.

Of the proteomic approaches, 2D PAGE is a well-known technique. Isolation of individual proteins from a sample such as serum is accomplished using sequential separation of proteins by isoelectric point and molecular weight. Typically, polypeptides are first separated by isoelectric point (the first dimension) and then separated by size using an electric current (the second dimension). In general, the second dimension is perpendicular to the first dimension. Because no two proteins with different sequences are identical on the basis of both size and charge, the result of 2D PAGE is a roughly square gel in which each protein occupies a unique spot. Analysis of the spots with chemical or antibody probes, or subsequent protein microsequencing can reveal the relative abundance of a given protein and the identity of the proteins in the sample.

Expression levels of a PSNA can be determined by any method known in the art, including PCR and other nucleic acid methods, such as ligase chain reaction (LCR) and nucleic acid sequence based amplification (NASBA), can be used to detect malignant cells for diagnosis and monitoring of various malignancies. For example, reverse-transcriptase PCR (RT-PCR) is a powerful technique which can be used to detect the presence of a specific mRNA population in a complex mixture of thousands of other mRNA species. In RT-PCR, an mRNA species is first reverse transcribed to complementary DNA (cDNA) with use of the enzyme reverse transcriptase; the cDNA is then amplified as in a standard PCR reaction.

Hybridization to specific DNA molecules (*e.g.*, oligonucleotides) arrayed on a solid support can be used to both detect the expression of and quantitate the level of expression of one or more PSNAs of interest. In this approach, all or a portion of one or more PSNAs is fixed to a substrate. A sample of interest, which may comprise RNA, *e.g.*, total RNA or polyA-selected mRNA, or a complementary DNA (cDNA) copy of the RNA is incubated with the solid support under conditions in which hybridization will occur between the DNA on the solid support and the nucleic acid molecules in the

sample of interest. Hybridization between the substrate-bound DNA and the nucleic acid molecules in the sample can be detected and quantitated by several means, including, without limitation, radioactive labeling or fluorescent labeling of the nucleic acid molecule or a secondary molecule designed to detect the hybrid.

5           The above tests can be carried out on samples derived from a variety of cells, bodily fluids and/or tissue extracts such as homogenates or solubilized tissue obtained from a patient. Tissue extracts are obtained routinely from tissue biopsy and autopsy material. Bodily fluids useful in the present invention include blood, urine, saliva or any other bodily secretion or derivative thereof. By blood it is meant to include whole blood,  
10       plasma, serum or any derivative of blood. In a preferred embodiment, the specimen tested for expression of PSNA or PSP includes, without limitation, prostate tissue, fluid obtained by bronchial alveolar lavage (BAL), sputum, prostate cells grown in cell culture, blood, serum, lymph node tissue and lymphatic fluid. In another preferred embodiment, especially when metastasis of a primary prostate cancer is known or  
15       suspected, specimens include, without limitation, tissues from brain, bone, bone marrow, liver, adrenal glands and colon. In general, the tissues may be sampled by biopsy, including, without limitation, needle biopsy, *e.g.*, transthoracic needle aspiration, cervical mediastinoscopy, endoscopic lymph node biopsy, video-assisted thoracoscopy, exploratory thoracotomy, bone marrow biopsy and bone marrow aspiration. *See Scott,*  
20       *supra* and Franklin, pp. 529-570, in Kane, *supra*. For early and inexpensive detection, assaying for changes in PSNAs or PSPs in cells in sputum samples may be particularly useful. Methods of obtaining and analyzing sputum samples is disclosed in Franklin, *supra*.

          All the methods of the present invention may optionally include determining the  
25       expression levels of one or more other cancer markers in addition to determining the expression level of a PSNA or PSP. In many cases, the use of another cancer marker will decrease the likelihood of false positives or false negatives. In one embodiment, the one or more other cancer markers include other PSNA or PSPs as disclosed herein. Other cancer markers useful in the present invention will depend on the cancer being tested and  
30       are known to those of skill in the art. In a preferred embodiment, at least one other cancer marker in addition to a particular PSNA or PSP is measured. In a more preferred



embodiment, at least two other additional cancer markers are used. In an even more preferred embodiment, at least three, more preferably at least five, even more preferably at least ten additional cancer markers are used.

### *Diagnosing*

5 In one aspect, the invention provides a method for determining the expression levels and/or structural alterations of one or more PSNAs and/or PSPs in a sample from a patient suspected of having prostate cancer. In general, the method comprises the steps of obtaining the sample from the patient, determining the expression level or structural alterations of a PSNA and/or PSP and then ascertaining whether the patient has prostate  
10 cancer from the expression level of the PSNA or PSP. In general, if high expression relative to a control of a PSNA or PSP is indicative of prostate cancer, a diagnostic assay is considered positive if the level of expression of the PSNA or PSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal  
15 human control. In contrast, if low expression relative to a control of a PSNA or PSP is indicative of prostate cancer, a diagnostic assay is considered positive if the level of expression of the PSNA or PSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control  
20 may be from a different patient or from uninvolved tissue of the same patient.

The present invention also provides a method of determining whether prostate cancer has metastasized in a patient. One may identify whether the prostate cancer has metastasized by measuring the expression levels and/or structural alterations of one or more PSNAs and/or PSPs in a variety of tissues. The presence of a PSNA or PSP in a  
25 certain tissue at levels higher than that of corresponding noncancerous tissue (*e.g.*, the same tissue from another individual) is indicative of metastasis if high level expression of a PSNA or PSP is associated with prostate cancer. Similarly, the presence of a PSNA or PSP in a tissue at levels lower than that of corresponding noncancerous tissue is indicative of metastasis if low level expression of a PSNA or PSP is associated with

prostate cancer. Further, the presence of a structurally altered PSNA or PSP that is associated with prostate cancer is also indicative of metastasis.

In general, if high expression relative to a control of a PSNA or PSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the PSNA or PSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a PSNA or PSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the PSNA or PSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control.

### *Staging*

The invention also provides a method of staging prostate cancer in a human patient. The method comprises identifying a human patient having prostate cancer and analyzing cells, tissues or bodily fluids from such human patient for expression levels and/or structural alterations of one or more PSNAs or PSPs. First, one or more tumors from a variety of patients are staged according to procedures well-known in the art, and the expression levels of one or more PSNAs or PSPs is determined for each stage to obtain a standard expression level for each PSNA and PSP. Then, the PSNA or PSP expression levels of the PSNA or PSP are determined in a biological sample from a patient whose stage of cancer is not known. The PSNA or PSP expression levels from the patient are then compared to the standard expression level. By comparing the expression level of the PSNAs and PSPs from the patient to the standard expression levels, one may determine the stage of the tumor. The same procedure may be followed using structural alterations of a PSNA or PSP to determine the stage of a prostate cancer.

### *Monitoring*

Further provided is a method of monitoring prostate cancer in a human patient. One may monitor a human patient to determine whether there has been metastasis and, if there has been, when metastasis began to occur. One may also monitor a human patient

to determine whether a preneoplastic lesion has become cancerous. One may also monitor a human patient to determine whether a therapy, *e.g.*, chemotherapy, radiotherapy or surgery, has decreased or eliminated the prostate cancer. The method comprises identifying a human patient that one wants to monitor for prostate cancer, periodically analyzing cells, tissues or bodily fluids from such human patient for expression levels of one or more PSNAs or PSPs, and comparing the PSNA or PSP levels over time to those PSNA or PSP expression levels obtained previously. Patients may also be monitored by measuring one or more structural alterations in a PSNA or PSP that are associated with prostate cancer.

10           If increased expression of a PSNA or PSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an increase in the expression level of a PSNA or PSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. One having ordinary skill in the art would recognize that if this were the case, then a  
15           decreased expression level would be indicative of no metastasis, effective therapy or failure to progress to a neoplastic lesion. If decreased expression of a PSNA or PSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an decrease in the expression level of a PSNA or PSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is  
20           cancerous, respectively. In a preferred embodiment, the levels of PSNAs or PSPs are determined from the same cell type, tissue or bodily fluid as prior patient samples. Monitoring a patient for onset of prostate cancer metastasis is periodic and preferably is done on a quarterly basis, but may be done more or less frequently.

          The methods described herein can further be utilized as prognostic assays to  
25           identify subjects having or at risk of developing a disease or disorder associated with increased or decreased expression levels of a PSNA and/or PSP. The present invention provides a method in which a test sample is obtained from a human patient and one or more PSNAs and/or PSPs are detected. The presence of higher (or lower) PSNA or PSP levels as compared to normal human controls is diagnostic for the human patient being at  
30           risk for developing cancer, particularly prostate cancer. The effectiveness of therapeutic agents to decrease (or increase) expression or activity of one or more PSNAs and/or PSPs

of the invention can also be monitored by analyzing levels of expression of the PSNAs and/or PSPs in a human patient in clinical trials or in *in vitro* screening assays such as in human cells. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the human patient or cells, as the case may be, to the agent  
5 being tested.

#### *Detection of Genetic Lesions or Mutations*

The methods of the present invention can also be used to detect genetic lesions or mutations in a PSG, thereby determining if a human with the genetic lesion is susceptible to developing prostate cancer or to determine what genetic lesions are responsible, or are  
10 partly responsible, for a person's existing prostate cancer. Genetic lesions can be detected, for example, by ascertaining the existence of a deletion, insertion and/or substitution of one or more nucleotides from the PSGs of this invention, a chromosomal rearrangement of PSG, an aberrant modification of PSG (such as of the methylation pattern of the genomic DNA), or allelic loss of a PSG. Methods to detect such lesions in  
15 the PSG of this invention are known to those having ordinary skill in the art following the teachings of the specification.

#### Methods of Detecting Noncancerous Prostate Diseases

In one aspect, the invention provides a method for determining the expression  
20 levels and/or structural alterations of one or more PSNAs and/or PSPs in a sample from a patient suspected of having or known to have a noncancerous prostate disease. In general, the method comprises the steps of obtaining a sample from the patient, determining the expression level or structural alterations of a PSNA and/or PSP, comparing the expression level or structural alteration of the PSNA or PSP to a normal  
25 prostate control, and then ascertaining whether the patient has a noncancerous prostate disease. In general, if high expression relative to a control of a PSNA or PSP is indicative of a particular noncancerous prostate disease, a diagnostic assay is considered positive if the level of expression of the PSNA or PSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times  
30 higher, than in preferably the same cells, tissues or bodily fluid of a normal human

control. In contrast, if low expression relative to a control of a PSNA or PSP is indicative of a noncancerous prostate disease, a diagnostic assay is considered positive if the level of expression of the PSNA or PSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

One having ordinary skill in the art may determine whether a PSNA and/or PSP is associated with a particular noncancerous prostate disease by obtaining prostate tissue from a patient having a noncancerous prostate disease of interest and determining which PSNAs and/or PSPs are expressed in the tissue at either a higher or a lower level than in normal prostate tissue. In another embodiment, one may determine whether a PSNA or PSP exhibits structural alterations in a particular noncancerous prostate disease state by obtaining prostate tissue from a patient having a noncancerous prostate disease of interest and determining the structural alterations in one or more PSNAs and/or PSPs relative to normal prostate tissue.

#### Methods for Identifying Prostate Tissue

In another aspect, the invention provides methods for identifying prostate tissue. These methods are particularly useful in, *e.g.*, forensic science, prostate cell differentiation and development, and in tissue engineering.

In one embodiment, the invention provides a method for determining whether a sample is prostate tissue or has prostate tissue-like characteristics. The method comprises the steps of providing a sample suspected of comprising prostate tissue or having prostate tissue-like characteristics, determining whether the sample expresses one or more PSNAs and/or PSPs, and, if the sample expresses one or more PSNAs and/or PSPs, concluding that the sample comprises prostate tissue. In a preferred embodiment, the PSNA encodes a polypeptide having an amino acid sequence selected from SEQ ID NO:23-31, or a homolog, allelic variant or fragment thereof. In a more preferred embodiment, the PSNA has a nucleotide sequence selected from SEQ ID NO:1-22, or a hybridizing nucleic acid, an allelic variant or a part thereof. Determining whether a

sample expresses a PSNA can be accomplished by any method known in the art.

Preferred methods include hybridization to microarrays, Northern blot hybridization, and quantitative or qualitative RT-PCR. In another preferred embodiment, the method can be practiced by determining whether a PSP is expressed. Determining whether a sample  
5 expresses a PSP can be accomplished by any method known in the art. Preferred methods include Western blot, ELISA, RIA and 2D PAGE. In one embodiment, the PSP has an amino acid sequence selected from SEQ ID NO:23-31, or a homolog, allelic variant or fragment thereof. In another preferred embodiment, the expression of at least two PSNAs and/or PSPs is determined. In a more preferred embodiment, the expression  
10 of at least three, more preferably four and even more preferably five PSNAs and/or PSPs are determined.

In one embodiment, the method can be used to determine whether an unknown tissue is prostate tissue. This is particularly useful in forensic science, in which small, damaged pieces of tissues that are not identifiable by microscopic or other means are  
15 recovered from a crime or accident scene. In another embodiment, the method can be used to determine whether a tissue is differentiating or developing into prostate tissue. This is important in monitoring the effects of the addition of various agents to cell or tissue culture, *e.g.*, in producing new prostate tissue by tissue engineering. These agents include, *e.g.*, growth and differentiation factors, extracellular matrix proteins and culture  
20 medium. Other factors that may be measured for effects on tissue development and differentiation include gene transfer into the cells or tissues, alterations in pH, aqueous: air interface and various other culture conditions.

#### Methods for Producing and Modifying Prostate Tissue

25 In another aspect, the invention provides methods for producing engineered prostate tissue or cells. In one embodiment, the method comprises the steps of providing cells, introducing a PSNA or a PSG into the cells, and growing the cells under conditions in which they exhibit one or more properties of prostate tissue cells. In a preferred embodiment, the cells are pluripotent. As is well-known in the art, normal prostate tissue  
30 comprises a large number of different cell types. Thus, in one embodiment, the engineered prostate tissue or cells comprises one of these cell types. In another

embodiment, the engineered prostate tissue or cells comprises more than one prostate cell type. Further, the culture conditions of the cells or tissue may require manipulation in order to achieve full differentiation and development of the prostate cell tissue. Methods for manipulating culture conditions are well-known in the art.

5           Nucleic acid molecules encoding one or more PSPs are introduced into cells, preferably pluripotent cells. In a preferred embodiment, the nucleic acid molecules encode PSPs having amino acid sequences selected from SEQ ID NO:23-31, or homologous proteins, analogs, allelic variants or fragments thereof. In a more preferred embodiment, the nucleic acid molecules have a nucleotide sequence selected from SEQ  
10 ID NO:1-22, or hybridizing nucleic acids, allelic variants or parts thereof. In another highly preferred embodiment, a PSG is introduced into the cells. Expression vectors and methods of introducing nucleic acid molecules into cells are well-known in the art and are described in detail, *supra*.

          Artificial prostate tissue may be used to treat patients who have lost some or all of  
15 their prostate function.

#### Pharmaceutical Compositions

          In another aspect, the invention provides pharmaceutical compositions comprising the nucleic acids, nucleic acid parts, polypeptides, protein fusions,  
20 polypeptide fragments, polypeptide analogs, derivatives and homologous polypeptides, antibodies, antibody derivatives, antibody fragments, agonists, antagonists, and inhibitors of the present invention. In a preferred embodiment, the pharmaceutical composition comprises a PSNA or part thereof. In a more preferred embodiment, the PSNA has a nucleotide sequence selected from the group consisting of SEQ ID NO:1-22, a nucleic  
25 acid that hybridizes thereto, an allelic variant thereof, or a nucleic acid that has substantial sequence identity thereto. In another preferred embodiment, the pharmaceutical composition comprises a PSP or fragment thereof. In a more preferred embodiment, the PSP having an amino acid sequence that is selected from the group consisting of SEQ ID NO:23-31, a polypeptide that is homologous thereto, a fusion  
30 protein comprising all or a portion of the polypeptide, or an analog or derivative thereof. In another preferred embodiment, the pharmaceutical composition comprises an anti-PSP

antibody, preferably an antibody that specifically binds to a PSP having an amino acid that is selected from the group consisting of SEQ ID NO:23-31, or an antibody that binds to a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof.

- 5           Such a composition typically contains from about 0.1 to 90% by weight of a therapeutic agent of the invention formulated in and/or with a pharmaceutically acceptable carrier or excipient.

Pharmaceutical formulation is a well-established art, and is further described in Gennaro (ed.), Remington: The Science and Practice of Pharmacy, 20<sup>th</sup> ed., Lippincott, 10 Williams & Wilkins (2000); Ansel *et al.*, Pharmaceutical Dosage Forms and Drug Delivery Systems, 7<sup>th</sup> ed., Lippincott Williams & Wilkins (1999); and Kibbe (ed.), Handbook of Pharmaceutical Excipients American Pharmaceutical Association, 3<sup>rd</sup> ed. (2000), the disclosures of which are incorporated herein by reference in their entireties, and thus need not be described in detail herein.

- 15           Briefly, formulation of the pharmaceutical compositions of the present invention will depend upon the route chosen for administration. The pharmaceutical compositions utilized in this invention can be administered by various routes including both enteral and parenteral routes, including oral, intravenous, intramuscular, subcutaneous, inhalation, topical, sublingual, rectal, intra-arterial, intramedullary, intrathecal, intraventricular, 20 transmucosal, transdermal, intranasal, intraperitoneal, intrapulmonary, and intrauterine.

Oral dosage forms can be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

- Solid formulations of the compositions for oral administration can contain suitable carriers or excipients, such as carbohydrate or protein fillers, such as sugars, 25 including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, or microcrystalline cellulose; gums including arabic and tragacanth; proteins such as gelatin and collagen; inorganics, such as kaolin, calcium carbonate, dicalcium phosphate, sodium chloride; and other agents such as acacia and 30 alginic acid.



Agents that facilitate disintegration and/or solubilization can be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate, microcrystalline cellulose, corn starch, sodium starch glycolate, and alginic acid.

5       Tablet binders that can be used include acacia, methylcellulose, sodium carboxymethylcellulose, polyvinylpyrrolidone (Povidone™), hydroxypropyl methylcellulose, sucrose, starch and ethylcellulose.

Lubricants that can be used include magnesium stearates, stearic acid, silicone fluid, talc, waxes, oils, and colloidal silica.

10       Fillers, agents that facilitate disintegration and/or solubilization, tablet binders and lubricants, including the aforementioned, can be used singly or in combination.

Solid oral dosage forms need not be uniform throughout. For example, dragee cores can be used in conjunction with suitable coatings, such as concentrated sugar solutions, which can also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures.

15       Oral dosage forms of the present invention include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds can be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

20       Additionally, dyestuffs or pigments can be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, *i.e.*, dosage.

25       Liquid formulations of the pharmaceutical compositions for oral (enteral) administration are prepared in water or other aqueous vehicles and can contain various suspending agents such as methylcellulose, alginates, tragacanth, pectin, kelgin, carrageenan, acacia, polyvinylpyrrolidone, and polyvinyl alcohol. The liquid formulations can also include solutions, emulsions, syrups and elixirs containing,

30

together with the active compound(s), wetting agents, sweeteners, and coloring and flavoring agents.

The pharmaceutical compositions of the present invention can also be formulated for parenteral administration. Formulations for parenteral administration can be in the  
5 form of aqueous or non-aqueous isotonic sterile injection solutions or suspensions.

For intravenous injection, water soluble versions of the compounds of the present invention are formulated in, or if provided as a lyophilate, mixed with, a physiologically acceptable fluid vehicle, such as 5% dextrose ("D5"), physiologically buffered saline, 0.9% saline, Hanks' solution, or Ringer's solution. Intravenous formulations may include  
10 carriers, excipients or stabilizers including, without limitation, calcium, human serum albumin, citrate, acetate, calcium chloride, carbonate, and other salts.

Intramuscular preparations, *e.g.* a sterile formulation of a suitable soluble salt form of the compounds of the present invention, can be dissolved and administered in a pharmaceutical excipient such as Water-for-Injection, 0.9% saline, or 5% glucose  
15 solution. Alternatively, a suitable insoluble form of the compound can be prepared and administered as a suspension in an aqueous base or a pharmaceutically acceptable oil base, such as an ester of a long chain fatty acid (*e.g.*, ethyl oleate), fatty oils such as sesame oil, triglycerides, or liposomes.

Parenteral formulations of the compositions can contain various carriers such as  
20 vegetable oils, dimethylacetamide, dimethylformamide, ethyl lactate, ethyl carbonate, isopropyl myristate, ethanol, polyols (glycerol, propylene glycol, liquid polyethylene glycol, and the like).

Aqueous injection suspensions can also contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran.  
25 Non-lipid polycationic amino polymers can also be used for delivery. Optionally, the suspension can also contain suitable stabilizers or agents that increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

Pharmaceutical compositions of the present invention can also be formulated to permit injectable, long-term, deposition. Injectable depot forms may be made by forming  
30 microencapsulated matrices of the compound in biodegradable polymers such as polylactide-polyglycolide. Depending upon the ratio of drug to polymer and the nature

of the particular polymer employed, the rate of drug release can be controlled. Examples of other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot injectable formulations are also prepared by entrapping the drug in microemulsions that are compatible with body tissues.

5           The pharmaceutical compositions of the present invention can be administered topically.

For topical use the compounds of the present invention can also be prepared in suitable forms to be applied to the skin, or mucus membranes of the nose and throat, and can take the form of lotions, creams, ointments, liquid sprays or inhalants, drops,  
10   tinctures, lozenges, or throat paints. Such topical formulations further can include chemical compounds such as dimethylsulfoxide (DMSO) to facilitate surface penetration of the active ingredient. In other transdermal formulations, typically in patch-delivered formulations, the pharmaceutically active compound is formulated with one or more skin penetrants, such as 2-N-methyl-pyrrolidone (NMP) or Azone. A topical semi-solid  
15   ointment formulation typically contains a concentration of the active ingredient from about 1 to 20%, *e.g.*, 5 to 10%, in a carrier such as a pharmaceutical cream base.

For application to the eyes or ears, the compounds of the present invention can be presented in liquid or semi-liquid form formulated in hydrophobic or hydrophilic bases as ointments, creams, lotions, paints or powders.

20           For rectal administration the compounds of the present invention can be administered in the form of suppositories admixed with conventional carriers such as cocoa butter, wax or other glyceride.

Inhalation formulations can also readily be formulated. For inhalation, various powder and liquid formulations can be prepared. For aerosol preparations, a sterile  
25   formulation of the compound or salt form of the compound may be used in inhalers, such as metered dose inhalers, and nebulizers. Aerosolized forms may be especially useful for treating respiratory disorders.

Alternatively, the compounds of the present invention can be in powder form for reconstitution in the appropriate pharmaceutically acceptable carrier at the time of  
30   delivery.

The pharmaceutically active compound in the pharmaceutical compositions of the present invention can be provided as the salt of a variety of acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms.

After pharmaceutical compositions have been prepared, they are packaged in an appropriate container and labeled for treatment of an indicated condition.

The active compound will be present in an amount effective to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

A "therapeutically effective dose" refers to that amount of active ingredient — for example PSP polypeptide, fusion protein, or fragments thereof, antibodies specific for PSP, agonists, antagonists or inhibitors of PSP — which ameliorates the signs or symptoms of the disease or prevents progression thereof; as would be understood in the medical arts, cure, although desired, is not required.

The therapeutically effective dose of the pharmaceutical agents of the present invention can be estimated initially by *in vitro* tests, such as cell culture assays, followed by assay in model animals, usually mice, rats, rabbits, dogs, or pigs. The animal model can also be used to determine an initial preferred concentration range and route of administration.

For example, the ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population) can be determined in one or more cell culture or animal model systems. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as LD50/ED50. Pharmaceutical compositions that exhibit large therapeutic indices are preferred.

The data obtained from cell culture assays and animal studies are used in formulating an initial dosage range for human use, and preferably provide a range of circulating concentrations that includes the ED50 with little or no toxicity. After administration, or between successive administrations, the circulating concentration of active agent varies within this range depending upon pharmacokinetic factors well-

known in the art, such as the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors specific to the subject requiring treatment. Factors that can be taken into account by the practitioner include the severity of the disease state, general health of the subject, age, weight, gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Where the therapeutic agent is a protein or antibody of the present invention, the therapeutic protein or antibody agent typically is administered at a daily dosage of 0.01 mg to 30 mg/kg of body weight of the patient (*e.g.*, 1mg/kg to 5 mg/kg). The pharmaceutical formulation can be administered in multiple doses per day, if desired, to achieve the total desired daily dose.

Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

Conventional methods, known to those of ordinary skill in the art of medicine, can be used to administer the pharmaceutical formulation(s) of the present invention to the patient. The pharmaceutical compositions of the present invention can be administered alone, or in combination with other therapeutic agents or interventions.

#### Therapeutic Methods

The present invention further provides methods of treating subjects having defects in a gene of the invention — *e.g.*, in expression, activity, distribution, localization, and/or solubility — which can manifest as a disorder of prostate function. As used herein, “treating” includes all medically-acceptable types of therapeutic intervention, including palliation and prophylaxis (prevention) of disease. The term

"treating" encompasses any improvement of a disease, including minor improvements. These methods are discussed below.

#### *Gene Therapy and Vaccines*

The isolated nucleic acids of the present invention can also be used to drive *in vivo* expression of the polypeptides of the present invention. *In vivo* expression can be driven from a vector — typically a viral vector, often a vector based upon a replication incompetent retrovirus, an adenovirus, or an adeno-associated virus (AAV) — for purpose of gene therapy. *In vivo* expression can also be driven from signals endogenous to the nucleic acid or from a vector, often a plasmid vector, such as pVAX1 (Invitrogen, Carlsbad, CA, USA), for purpose of “naked” nucleic acid vaccination, as further described in United States Patent Nos. 5,589,466; 5,679,647; 5,804,566; 5,830,877; 5,843,913; 5,880,104; 5,958,891; 5,985,847; 6,017,897; 6,110,898; 6,204,250, the disclosures of which are incorporated herein by reference in their entireties. For cancer therapy, it is preferred that the vector also be tumor-selective. *See, e.g., Doronin et al., J. Virol.* 75: 3314-24 (2001).

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a nucleic acid of the present invention is administered. The nucleic acid can be delivered in a vector that drives expression of a PSP, fusion protein, or fragment thereof, or without such vector. Nucleic acid compositions that can drive expression of a PSP are administered, for example, to complement a deficiency in the native PSP, or as DNA vaccines. Expression vectors derived from virus, replication deficient retroviruses, adenovirus, adeno-associated (AAV) virus, herpes virus, or vaccinia virus can be used as can plasmids. *See, e.g., Cid-Arregui, supra.* In a preferred embodiment, the nucleic acid molecule encodes a PSP having the amino acid sequence of SEQ ID NO:23-31, or a fragment, fusion protein, allelic variant or homolog thereof.

In still other therapeutic methods of the present invention, pharmaceutical compositions comprising host cells that express a PSP, fusions, or fragments thereof can be administered. In such cases, the cells are typically autologous, so as to circumvent xenogeneic or allotypic rejection, and are administered to complement defects in PSP

production or activity. In a preferred embodiment, the nucleic acid molecules in the cells encode a PSP having the amino acid sequence of SEQ ID NO:23-31, or a fragment, fusion protein, allelic variant or homolog thereof.

#### *Antisense Administration*

5 Antisense nucleic acid compositions, or vectors that drive expression of a PSG antisense nucleic acid, are administered to downregulate transcription and/or translation of a PSG in circumstances in which excessive production, or production of aberrant protein, is the pathophysiologic basis of disease.

Antisense compositions useful in therapy can have a sequence that is  
10 complementary to coding or to noncoding regions of a PSG. For example, oligonucleotides derived from the transcription initiation site, *e.g.*, between positions -10 and +10 from the start site, are preferred.

Catalytic antisense compositions, such as ribozymes, that are capable of sequence-specific hybridization to PSG transcripts, are also useful in therapy. *See, e.g.*,  
15 Phylactou, *Adv. Drug Deliv. Rev.* 44(2-3): 97-108 (2000); Phylactou *et al.*, *Hum. Mol. Genet.* 7(10): 1649-53 (1998); Rossi, *Ciba Found. Symp.* 209: 195-204 (1997); and Sigurdsson *et al.*, *Trends Biotechnol.* 13(8): 286-9 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Other nucleic acids useful in the therapeutic methods of the present invention are  
20 those that are capable of triplex helix formation in or near the PSG genomic locus. Such triplexing oligonucleotides are able to inhibit transcription. *See, e.g.*, Intody *et al.*, *Nucleic Acids Res.* 28(21): 4283-90 (2000); McGuffie *et al.*, *Cancer Res.* 60(14): 3790-9 (2000), the disclosures of which are incorporated herein by reference. Pharmaceutical compositions comprising such triplex forming oligos (TFOs) are administered in  
25 circumstances in which excessive production, or production of aberrant protein, is a pathophysiologic basis of disease.

In a preferred embodiment, the antisense molecule is derived from a nucleic acid molecule encoding a PSP, preferably a PSP comprising an amino acid sequence of SEQ ID NO:23-31, or a fragment, allelic variant or homolog thereof. In a more preferred  
30 embodiment, the antisense molecule is derived from a nucleic acid molecule having a

nucleotide sequence of SEQ ID NO:1-22, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

*Polypeptide Administration*

In one embodiment of the therapeutic methods of the present invention, a  
5 therapeutically effective amount of a pharmaceutical composition comprising a PSP, a fusion protein, fragment, analog or derivative thereof is administered to a subject with a clinically-significant PSP defect.

Protein compositions are administered, for example, to complement a deficiency in native PSP. In other embodiments, protein compositions are administered as a vaccine  
10 to elicit a humoral and/or cellular immune response to PSP. The immune response can be used to modulate activity of PSP or, depending on the immunogen, to immunize against aberrant or aberrantly expressed forms, such as mutant or inappropriately expressed isoforms. In yet other embodiments, protein fusions having a toxic moiety are administered to ablate cells that aberrantly accumulate PSP.

15 In a preferred embodiment, the polypeptide is a PSP comprising an amino acid sequence of SEQ ID NO:23-31, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the polypeptide is encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO:1-22, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

20 *Antibody, Agonist and Antagonist Administration*

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an antibody (including fragment or derivative thereof) of the present invention is administered. As is well-known, antibody compositions are administered, for example,  
25 to antagonize activity of PSP, or to target therapeutic agents to sites of PSP presence and/or accumulation. In a preferred embodiment, the antibody specifically binds to a PSP comprising an amino acid sequence of SEQ ID NO:23-31, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antibody specifically binds to a PSP encoded by a nucleic acid molecule having a



nucleotide sequence of SEQ ID NO:1-22, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

The present invention also provides methods for identifying modulators which bind to a PSP or have a modulatory effect on the expression or activity of a PSP.

5 Modulators which decrease the expression or activity of PSP (antagonists) are believed to be useful in treating prostate cancer. Such screening assays are known to those of skill in the art and include, without limitation, cell-based assays and cell-free assays. Small molecules predicted via computer imaging to specifically bind to regions of a PSP can also be designed, synthesized and tested for use in the imaging and treatment of prostate  
10 cancer. Further, libraries of molecules can be screened for potential anticancer agents by assessing the ability of the molecule to bind to the PSPs identified herein. Molecules identified in the library as being capable of binding to a PSP are key candidates for further evaluation for use in the treatment of prostate cancer. In a preferred embodiment, these molecules will downregulate expression and/or activity of a PSP in cells.

15 In another embodiment of the therapeutic methods of the present invention, a pharmaceutical composition comprising a non-antibody antagonist of PSP is administered. Antagonists of PSP can be produced using methods generally known in the art. In particular, purified PSP can be used to screen libraries of pharmaceutical agents, often combinatorial libraries of small molecules, to identify those that specifically  
20 bind and antagonize at least one activity of a PSP.

In other embodiments a pharmaceutical composition comprising an agonist of a PSP is administered. Agonists can be identified using methods analogous to those used to identify antagonists.

In a preferred embodiment, the antagonist or agonist specifically binds to and  
25 antagonizes or agonizes, respectively, a PSP comprising an amino acid sequence of SEQ ID NO:23-31, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a PSP encoded by a nucleic acid molecule having a  
30 nucleotide sequence of SEQ ID NO:1-22, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

*Targeting Prostate Tissue*

The invention also provides a method in which a polypeptide of the invention, or an antibody thereto, is linked to a therapeutic agent such that it can be delivered to the prostate or to specific cells in the prostate. In a preferred embodiment, an anti-PSP antibody is linked to a therapeutic agent and is administered to a patient in need of such  
5 therapeutic agent. The therapeutic agent may be a toxin, if prostate tissue needs to be selectively destroyed. This would be useful for targeting and killing prostate cancer cells. In another embodiment, the therapeutic agent may be a growth or differentiation factor, which would be useful for promoting prostate cell function.

In another embodiment, an anti-PSP antibody may be linked to an imaging agent  
10 that can be detected using, *e.g.*, magnetic resonance imaging, CT or PET. This would be useful for determining and monitoring prostate function, identifying prostate cancer tumors, and identifying noncancerous prostate diseases.

## EXAMPLES

The present invention is further described by the following examples. The  
15 examples are provided solely to illustrate the invention by reference to specific embodiments. These exemplifications, while illustrating certain aspects of the invention, do not portray the limitations or circumscribe the scope of the disclosed invention.

All examples outlined here were carried out using standard techniques, which are well known and routine to those of skill in the art, except where otherwise described in detail.  
20 Routine molecular biology techniques of the following example can be carried out as described in standard laboratory manuals, such as Sambrook et al., MOLECULAR CLONING: A LABORATORY MANUAL, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989).

### 25 EXAMPLE 1: Relative Quantitation of Gene Expression

Real-Time quantitative PCR with fluorescent Taqman probes is a quantitation detection system utilizing the 5'- 3' nuclease activity of Taq DNA polymerase. The method uses an internal fluorescent oligonucleotide probe (Taqman) labeled with a 5' reporter dye and a downstream, 3' quencher dye. During PCR, the 5'-3' nuclease activity  
30 of Taq DNA polymerase releases the reporter, whose fluorescence can then be detected

by the laser detector of the Model 7700 Sequence Detection System (PE Applied Biosystems, Foster City, CA, USA).

Amplification of an endogenous control is used to standardize the amount of sample RNA added to the reaction and normalize for Reverse Transcriptase (RT) efficiency. Either cyclophilin, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) or 18S ribosomal RNA (rRNA) was used as this endogenous control. To calculate relative quantitation between all the samples studied, the target RNA levels for one sample was used as the basis for comparative results (calibrator). Quantitation relative to the "calibrator" can be obtained using the standard curve method or the comparative method (User Bulletin #2: ABI PRISM 7700 Sequence Detection System).

The tissue distribution and the level of the target gene were determined for each sample of normal and cancer tissue. Total RNA was extracted from normal tissues, cancer tissues and from cancers and the corresponding matched adjacent tissues. Subsequently, first strand cDNA was prepared with reverse transcriptase and the polymerase chain reaction was done using primers and Taqman probe specific to each target gene. The results were analyzed using the ABI PRISM 7700 Sequence Detector. The absolute numbers are relative levels of expression of the target gene in a particular tissue compared to the calibrator tissue.

Table 1 provides a list of the Prostate Specific Nucleic Acids of the present invention.

Sequences	Gene ID	dDx Code
1	14007	Pro130
2	480173	Pro122
3	216181	Pro123
4	897942	Pro132
5	29050	Pro133
6	475721	SQPro003
7	66398	SQPro004
8	898372	SQPro006
9	441230	SQPro007
10	205010	SQPro008
11	399540	SQPro009
12	902736	SQPro010
13	903252	SQPro011
14	410181	SQPro012

15	475299.1	SQPro038
16	221059.14	SQPro039
17	408290.1	SQPro040
18	346369.1	SQPro041
19	221011.3	SQPro042
20	214783.1	SQPro043
21	240421.1	SQPro044
22	154838.1	SQPro045

**SEQ ID NO:1 (Pro130)**

- 5 Table 2 provides the absolute numbers which are relative levels of expression of Pro130 in 24 normal different tissues. All the values are compared to normal prostate (calibrator). These RNA samples are commercially pools, originated by pooling samples of a particular tissue from different individuals.

**TABLE 2: Pro130**

Tissue	NORMAL
Adrenal Gland	0.00
Bladder	0.00
Brain	0.00
Cervix	0.00
Colon	0.00
Endometrium	0.00
Esophagus	0.00
Heart	0.00
Kidney	0.00
Liver	0.00
Lung	0.00
Mammary Gland	0.00
Muscle	0.00
Ovary	0.00
Pancreas	0.00
Prostate	1.00
Rectum	0.00
Small Intestine	0.00
Spleen	0.00
Stomach	0.00
Testis	0.00
Thymus	0.00
Trachea	0.00
Uterus	0.00

10 0=negative

The relative levels of expression in Table 2 show that Pro130 mRNA expression is high in prostate (1.0) compared with all other normal tissues analyzed. All other tissues analyzed show undetectable Pro130 mRNA expression.

The absolute numbers in Table 2 were obtained analyzing pools of samples of a particular tissue from different individuals. They can not be compared to the absolute numbers originated from RNA obtained from tissue samples of a single individual in Table 3.

Table 3 shows the absolute numbers which are relative levels of expression of Pro130 in 52 pairs of matching samples and 3 prostate normal, and 17 prostatitis and benign prostate hyperplasia samples. All the values are compared to normal prostate (calibrator). A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

15 **Table 3: Pro130**

Sample ID	Tissue	CANCER	PROSTATITIS & BENIGN HYPERPLASIA	MATCHING NORMAL ADJACENT	NORMAL
Pro73P	Prostate 1				0.05
Pro77P	Prostate 2				0.27
ProC153	Prostate 3				0.00
Pro53P	Prostate 4				0.23
Pro101XB	Prostate 5	4.21		0.48	
Pro109XB	Prostate 6	0.01		0.07	
Pro125XB	Prostate 7	0.04		0.06	
Pro12B	Prostate 8	1.54		0.05	
Pro13XB	Prostate 9	0.01		0.06	
Pro23B	Prostate 10	0.61		0.43	
Pro110	Prostate 11	0.01		0.13	
Pro326	Prostate 12	0.04		0.07	
Pro34B	Prostate 13	0.61		0.31	
Pro65XB	Prostate 14	0.44		2.39	
Pro69XB	Prostate 15	0.02		0.01	
Pro78XB	Prostate 16	0.60		0.46	
Pro84XB	Prostate 17	0.67		0.05	
Pro90XB	Prostate 18	0.37		0.28	
Pro91XB	Prostate 19	1.36		0.31	
Pro20XB	Prostate 20	0.05		0.00	
ProC215	Prostate 21	1.34		0.00	
ProC234	Prostate 22	0.38		0.00	
ProC280	Prostate 23	0.68		0.00	
Pro588P	Prostate 24	0.22		0.00	

Pro10R	Prostate 25 (prostatitis)		0.02		
Pro20R	Prostate 26 (prostatitis)		0.09		
Pro10P	Prostate 27 (BPH)		0.01		
Pro13P	Prostate 28 (BPH)		0.00		
Pro258	Prostate 29 (BPH)		0.18		
Pro263C	Prostate 30 (BPH)		0.2		
Pro267A	Prostate 31 (BPH)		0.00		
Pro271A	Prostate 32 (BPH)		0.01		
Pro460Z	Prostate 33 (BPH)		0.50		
Pro65P	Prostate 34 (BPH)		0.01		
Pro705P	Prostate 35 (BPH)		0.01		
Pro784P	Prostate 36 (BPH)		0.01		
Pro83P	Prostate 37 (BPH)		0.01		
Pro855P	Prostate 38 (BPH)		0.13		
ProC003P	Prostate 39 (BPH)		0.03		
ProC032	Prostate 40 (BPH)		0.50		
ProC034P	Prostate 41 (BPH)		0.00		
Testis 39X	Testis 1	0.00		0.00	
Testis 647T	Testis 2	0.03		0.00	
Testis 663T	Testis 3	0.00		0.00	
Bladder 32XK	Bladder 1	0.00		0.00	
Bladder 46XK	Bladder 2	0.01		0.00	
Bladder 66X	Bladder 3	0.00		0.00	
Bladder TR14	Bladder 4	0.00		0.00	
Bladder TR17	Bladder 5	0.00		0.00	
Kidney 10.006XD	Kidney 1	0.00		0.00	
Kidney 10.007XD	Kidney 2	0.00		0.00	
Kidney 10.009XD	Kidney 3	0.00		0.00	
Kidney 10.00XD	Kidney 4	0.00		0.00	
Kidney 11XD	Kidney 5	0.00		0.00	
Kidney 124D	Kidney 6	0.00		0.00	

Liver 15XA	Liver 1	0.00		0.00	
Liver 174L	Liver 2	0.00		0.00	
Lung 143L	Lung 1	0.00		0.00	
Lung 223L	Lung 2	0.00		0.00	
Colon 132C	Colon 1	0.00		0.00	
Colon AC19	Colon 2	0.00		0.00	
Colon AS12	Colon 3	0.00		0.00	
Mammary 162X	Mammary 1	0.00		0.00	
Mammary 19DN	Mammary 2	0.00		0.00	
Ovary A0.0082	Ovary 1	0.00		0.00	
Ovary A0.0084	Ovary 2	0.00		0.00	
Ovary 10.003X	Ovary 3	0.00		0.00	
Endometrium 10.00479	Endometrium 1	0.00		0.00	
Endometrium 12XA	Endometrium 2	0.00		0.00	
Endometrium 28XA	Endometrium 3	0.00		0.00	
Uterus 135XO	Uterus 1	0.00		0.00	
Uterus 141XO	Uterus 2	0.00		0.00	

0.00= Negative

In the analysis of matching samples, higher expression of Pro130 is detected in prostate samples showing a high degree of tissue specificity for prostate tissue. These results confirm the tissue specificity results obtained with normal pooled samples (Table 3).

Furthermore, we compared the level of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual. This comparison provides an indication of specificity for the cancer stage (e.g. higher levels of mRNA expression in the cancer sample compared to the normal adjacent). Table 3 shows overexpression of Pro130 in 40% of the prostate matching samples tested (6 out of total of 15 prostate matching samples).

Altogether, the high level of tissue specificity, plus the mRNA differential expression in the prostate matching samples tested are believed to make Pro130 a good marker for diagnosing, monitoring, staging, imaging and treating prostate cancer.

#### Primers Used for QPCR Expression Analysis

Forward primer:

GGAGGCAGAAGAGGAAGTCAGA (SEQ ID NO:32)

Reverse primer:

GCCATCCATGTTTCTCAGTTCC (SEQ ID NO:33)

Probe:

5 ATGTGCTGTGAAATCTACTACCGTTTGCTGG (SEQ ID NO:34)

#### SEQ ID NO: 2 Pro122

Table 4 shows the absolute numbers which are relative levels of expression of Pro122 in 12 normal different tissues. All the values are compared to normal endometrium (calibrator). These RNA samples are commercially pools, originated by pooling samples of a particular tissue from different individuals.

**Table 4: Pro122**

Tissue	NORMAL
Colon	0.01
Endometrium	1.00
Kidney	0.00
Liver	0.00
Ovary	0.03
Pancreas	0.03
Prostate	79.34
Small Intestine	0.00
Spleen	0.05
Stomach	0.03
Testis	12.42
Uterus	0.00

15 0=negative

The relative levels of expression in Table 4 show that Pro122 mRNA expression is high in prostate compared with all other normal tissues analyzed.

The absolute numbers in Table 4 were obtained analyzing pools of samples of a particular tissue from different individuals. They can not be compared to the absolute numbers originated from RNA obtained from tissue samples of a single individual in Table 5.

Table 5 shows the absolute numbers which are relative levels of expression of Pro122 in matching samples and some unmatched samples. All the values are compared to normal endometrium (calibrator). A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.



Table 5: Pro122

Sample ID	Tissue	CANCER	PROSTATITIS & BENIGN HYPERPLASIA	MATCHING NORMAL ADJACENT	NORMAL
Pro73P	Prostate 1				16.62
Pro77P	Prostate 2				22.32
ProC153	Prostate 3				2.03
Pro53P	Prostate 4				237.21
Pro101XB	Prostate 5	206.85		274.09	
Pro109XB	Prostate 6	5.30		37.85	
Pro110	Prostate 7	6.02		88.24	
Pro125XB	Prostate 8	13.69		21.62	
Pro12B	Prostate 9	68.20		1.65	
Pro13XB	Prostate 10	1.41		30.38	
Pro23B	Prostate 11	263.20		127.56	
Pro78XB	Prostate 12	21.33		17.27	
Pro90XB	Prostate 13	1.92		40.09	
Pro326	Prostate 14	0.00		0.00	
Pro34B	Prostate 15	0.00		0.00	
Pro65XB	Prostate 16	23.59		41.64	
Pro69XB	Prostate 17	9.71		12.25	
Pro84XB	Prostate 18	263.20		17.51	
Pro91X	Prostate 19	11.79		34.42	
Pro10R	Prostate 20 (prostatitis )		8.34		
Pro20R	Prostate 21 (prostatitis )		19.16		
Pro258	Prostate 21 (BPH)		2.49		
Pro263C	Prostate 22 (BPH)		20.04		
Pro267A	Prostate 23 (BPH)		3.67		
Pro271A	Prostate 24 (BPH)		7.09		
Pro10P	Prostate 25 (BPH)		46.69		
Pro13P	Prostate 26 (BPH)		1.56		
Pro277P	Prostate 27 (BPH)		54.38		
Pro34P	Prostate 28 (BPH)		25.81		
Pro460Z	Prostate 29 (BPH)		13.74		
Pro65P	Prostate 30 (BPH)		5.74		
Pro705P	Prostate 31 (BPH)		0.69		

Pro784P	Prostate 32 (BPH)		12.91		
Pro83P	Prostate 33 (BPH)		2.17		
Pro855P	Prostate 34 (BPH)		78.79		
ProC003P	Prostate 35 (BPH)		0.00		
ProC032	Prostate 36 (BPH)		0.23		
ProC034P	Prostate 38 (BPH)		13.45		
Testis 647T	Testis 1	0.00		0.72	
Testis 663T	Testis 2	0.00		1.84	
Testis 39X	Testis 3	0.00		2.35	
Bladder 32XK	Bladder 1	0.01		0.00	
Bladder 46XK	Bladder 2	0.00		0.01	
Bladder 66X	Bladder 3	0.00		0.00	
Bladder TR14	Bladder 4	0.00		0.00	
Kidney 106XD	Kidney 1	0.22		0.00	
Kidney 107XD	Kidney 2	0.16		0.16	
Kidney 109XD	Kidney 3	0.15		0.03	
Kidney 10XD	Kidney 4	0.09		0.28	
Lung 143L	Lung 1	0.00		0.00	
Lung 205L	Lung 2	0.00		0.00	
Pancreas 71XL	Pancreas 1	0.00		0.00	
Pancreas 77X	Pancreas 2	0.00		0.00	
Pancreas 82XP	Pancreas 3	0.02		0.00	
Colon AC19	Colon 1	0.08		0.00	
Colon AS12	Colon 2	0.00		0.00	
Colon AS43	Colon 3	0.00		0.00	
Mammary 162X	Mammary 1	0.00		0.00	
Mammary 173M	Mammary 2	0.00		0.00	
Mammary 12B	Mammary 3	0.00			
Mammary 12X	Mammary 4	0.00			
Ovary 988Z	Ovary 1	1.32			
Ovary 9RA	Ovary 2			0.00	
Ovary A084	Ovary 3	0.00		0.00	
Ovary G010	Ovary 4	0.08		0.02	
Ovary G021	Ovary 5	0.01		0.00	
Ovr 10050	Ovary 6	16.86			
Ovr 1028	Ovary 7	0.00			
Ovr 18GA	Ovary 8			0.05	

Ovr 206I	Ovary 9			0.08	
Ovary 103X	Ovary 10	0.00		0.00	
Ovary 638A	Ovary 11	0.00			
Ovary 638O	Ovary 12			0.00	
Ovary A082	Ovary 13	0.00		0.00	
Ovary 3710	Ovary 14	0.00			
Ovary 35GA	Ovary 15			0.00	
Ovary 63A	Ovary 16	0.00			
Ovary C360	Ovary 17	0.00			
Ovary 50GB	Ovary 18			0.00	
Cervix KS52	Cervix 1	0.01		0.24	
Cervix KS83	Cervix 2	0.00		0.27	
Endo 10479	Endometrium 1	0.10		0.00	
Endo 12XA	Endometrium 2	0.21		0.00	
Endometrium 28XA	Endometrium 3	0.02		0.01	
Endometrium 3AX	Endometrium 4	0.95		0.15	
Uterus 135XO	Uterus 1	0.11		0.02	
Uterus 141XO	Uterus 2	0.18		0.00	
Uterus 23XU	Uterus 3	0.07		0.00	

0.00= Negative

In the analysis of matching samples, higher expression of Pro122 is detected in prostate samples showing a high degree of tissue specificity for prostate tissue. These results confirm the tissue specificity results obtained with normal pooled samples (Table 4).

Furthermore, we compared the level of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual. This comparison provides an indication of specificity for the cancer stage (e.g. higher levels of mRNA expression in the cancer sample compared to the normal adjacent). Table 5 shows overexpression of Pro122 in 20% of the prostate matching samples tested (3 out of total of 15 prostate matching samples).

Altogether, the high level of tissue specificity, plus the mRNA differential expression in the prostate matching samples tested are believed to make Pro122 a good marker for diagnosing, monitoring, staging, imaging and treating prostate cancer.

**Primers Used for QPCR Expression Analysis**

Forward primer:

GGTGGCATGTGTTTTACTTTCA (SEQ ID NO:35)

Reverse primer:

5 AACGGGTCTTAATAATCAAATGACTC (SEQ ID NO:36)

Probe:

AATTTCCATCATTCTAACAAAATCAATTTTTCAGA (SEQ ID NO:37)

**10 SEQ ID NO:3 (Pro123)**

Table 6 shows the absolute numbers which are relative levels of expression of Pro123 in 12 normal different tissues. All the values are compared to normal stomach (calibrator). These RNA samples are commercially pools, originated by pooling samples of a particular tissue from different individuals.

**Table 6: Pro 123**

Tissue	NORMAL
Colon	0.68
Endometrium	1.64
Kidney	0.06
Liver	0.01
Ovary	0.08
Pancreas	0.02
Prostate	2.51
Small Intestine	1.17
Spleen	0.01
Stomach	1.00
Testis	0.45
Uterus	6.61

0=negative

20 The relative levels of expression in Table 6 show that Pro123 mRNA expression is high in prostate compared with all other normal tissues analyzed.

The absolute numbers in Table 6 were obtained analyzing pools of samples of a particular tissue from different individuals. They can not be compared to the absolute numbers originated from RNA obtained from tissue samples of a single individual in

25 Table 7.

Table 7 shows the absolute numbers which are relative levels of expression of Pro123 in matching samples and some unmatched samples. All the values are compared to normal stomach (calibrator). A matching pair is formed by mRNA from the cancer

sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

Table 7: Pro123

5

Sample ID	Tissue	CANCER	PROSTATITIS & BENIGN HYPERPLASIA	MATCHING NORMAL ADJACENT	NORMAL
Pro53P	Prostate 1				0.92
Pro73P	Prostate 2				0.84
Pro77P	Prostate 3				0.29
Pro101XB	Prostate 4	1.00		0.90	
Pro109XB	Prostate 5	14.50		1.10	
Pro110	Prostate 6	0.90		0.90	
Pro125XB	Prostate 7	9.60		2.30	
Pro12B	Prostate 8	1.80		0.20	
Pro13XB	Prostate 9	0.58		2.20	
Pro23B	Prostate 10	1.07		0.54	
Pro34B	Prostate 11	0.95		1.08	
Pro65XB	Prostate 12	7.01		3.53	
Pro69XB	Prostate 13	18.00		0.84	
Pro10R	Prostate 14 (prostatitis)		0.12		
Pro20R	Prostate 15 (prostatitis)		0.52		
Pro258	Prostate 16 (BPH)		0.66		
Pro263C	Prostate 17 (BPH)		0.26		
Pro267A	Prostate 18 (BPH)		0.09		
Pro271A	Prostate 19 (BPH)		0.12		
Pro13P	Prostate 20 (BPH)		0.90		
Pro377P	Prostate 21 (BPH)		0.48		
Testis 39X	Testis 1	0.30		0.80	
Testis 647T	Testis 2	0.21		0.30	
Kidney 106XD	Kidney 1	0.00		0.40	
Kidney 107XD	Kidney 2	1.80		0.10	
Bladder 32XK	Bladder 1	0.00		0.10	
Bladder 46XK	Bladder 2	1.40		0.00	
Bladder TR14	Bladder 4	0.50		0.20	
Pancreas 71XL	Pancreas 1	0.00		0.10	
Pancreas 77X	Pancreas 2	0.20		0.00	
Pancreas 82XP	Pancreas 3	0.05		1.03	
Stomach 115S	Stomach 1	0.70		2.10	

Colon AC19	Colon 1	0.10		1.10	
Colon AS12	Colon 2	0.10		0.40	
Colon AS43	Colon 3	0.76		0.12	
Small Intestine 21XA	Small Intestine 1	0.12		0.46	
Small Intestine H89	Small Intestine 2	1.10		3.40	
Endo 10479	Endometrium 1	0.40		1.40	
Endo 12XA	Endometrium 2	0.09		0.50	
Endo 28XA	Endometrium 3	0.17		0.08	
Endo 3AX	Endometrium 4	1.60			
Ovr 10050	Ovary 1	0.20			
Ovr 1028	Ovary 2			0.20	
Ovr 18GA	Ovary 3			0.40	
Ovr 206I	Ovary 4	1.47		0.71	
Uterus 135XO	Uterus 1	2.27		4.23	
Uterus 141XO	Uterus 2	0.20		1.89	
Uterus 23XU	Uterus 3	0.30		0.80	

0.00= Negative

In the analysis of matching samples, higher expression of Pro123 is detected in prostate samples showing a high degree of tissue specificity for prostate tissue. These results confirm the tissue specificity results obtained with normal pooled samples (Table 6).

Furthermore, we compared the level of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual. This comparison provides an indication of specificity for the cancer stage (e.g. higher levels of mRNA expression in the cancer sample compared to the normal adjacent). Table 7 shows overexpression of Pro123 in 60% of the prostate matching samples tested (6 out of total of 10 prostate matching samples).

Altogether, the high level of tissue specificity, plus the mRNA differential expression in the prostate matching samples tested are believed to make Pro123 a good marker for diagnosing, monitoring, staging, imaging and treating prostate cancer.

#### Primers Used for QPCR Expression Analysis

Forward primer:

GGTCGTATGTATTTCCAGGTGAG (SEQ ID NO:38)

Reverse primer:

TTGTCATCTTGCTGTTTCTAGTGAT (SEQ ID NO:39)

Probe:

AGGCTGCTGACTTTACCATCTGAGGC (SEQ ID NO:40)

# 5 SEQ ID NO:4 (Pro132)

Table 8 shows the absolute numbers are relative levels of expression of Pro132 in 24 normal different tissues. All the values are compared to normal rectum (calibrator). These RNA samples are commercially pools, originated by pooling samples of a particular tissue from different individuals.

**Table 8: Pro132**

Tissue	NORMAL
Adrenal Gland	0.27
Bladder	0.00
Brain	1.19
Cervix	0.17
Colon	0.08
Endometrium	2.73
Esophagus	0.07
Heart	0.05
Kidney	0.28
Liver	0.11
Lung	3.18
Mammary Gland	0.79
Muscle	0.08
Ovary	1.43
Pancreas	4.87
Prostate	10.20
Rectum	1.00
Small Intestine	0.34
Spleen	9.48
Stomach	0.62
Testis	0.55
Thymus	11.63
Trachea	2.39
Uterus	1.05

0=negative

15 The relative levels of expression in Table 8 show that Pro132 mRNA expression is high in prostate (1.0) compared with all other normal tissues analyzed.

The absolute numbers in Table 8 were obtained analyzing pools of samples of a particular tissue from different individuals. They can not be compared to the absolute numbers originated from RNA obtained from tissue samples of a single individual in Table 9.

- 5 Table 9 shows the absolute numbers which are relative levels of expression of Pro132 in matching samples and some unmatched samples. All the values are compared to normal rectum (calibrator). A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

10

**Table 9: Pro132**

Sample ID	Tissue	CANCER	PROSTATITIS & BENIGN HYPERPLASIA	MATCHING NORMAL ADJACENT	NORMAL
Pro73P	Prostate 1				1.55
Pro77P	Prostate 2				1.48
Pro101XB	Prostate 3	19.49		3.94	
Pro65XB	Prostate 4	5.56		3.28	
Pro23B	Prostate 5	8.75		3.20	
Pro90XB	Prostate 6	1.26		2.87	
Pro91XB	Prostate 7	2.97		3.11	
Pro125XB	Prostate 8	2.21		1.86	
Pro12B	Prostate 9	10.09		0.55	
Pro69XB	Prostate 10	3.07		2.08	
Pro109XB	Prostate 11	2.27		3.04	
Pro110	Prostate 12	1.15		4.92	
Pro13XB	Prostate 13	0.65		4.48	
Pro34B	Prostate 14	16.11		7.11	
Pro78XB	Prostate 15	3.26		9.85	
Pro84XB	Prostate 16	1.89		1.34	
Pro10R	Prostate 17 (prostatitis)		0.48		
Pro20R	Prostate 18 (prostatitis)		1.42		
Pro10P	Prostate 19 (BPH)		1.21		
Pro13P	Prostate 20 (BPH)		0.25		
Pro34P	Prostate 21 (BPH)		1.17		
Pro277P	Prostate 22 (BPH)		3.10		
Kidney 716K	Kidney 1	0.91		0.13	
Kidney 106XD	Kidney 2	1.57		0.11	



Kidney 107XD	Kidney 3	0.64		0.40	
Bladder 32XK	Bladder 1	1.28		0.07	
Bladder 46XK	Bladder 2	0.00		0.07	
Pancreas 77X	Pancreas 1	2.06		0.66	
Pancreas 82XP	Pancreas 2	0.97		1.85	
Pancreas 92X	Pancreas 3	2.80		1.45	
Stomach MT54	Stomach 1	0.84		0.52	
Stomach 88S	Stomach 2	0.23		0.75	
Stomach 915S	Stomach 3	0.00		0.62	
Colon B56	Colon 1	1.26		0.33	
Colon DC22	Colon 2	0.15		0.16	
Colon C9XR	Colon 3	0.19		0.38	
Lung LC80	Lung 1	5.06		0.42	
Lung 143L	Lung 2	0.58		0.14	
Lung 205L	Lung 3	3.05		0.00	
Endometrium 5XA	Endometrium 1	0.46		0.60	
Endometrium 65RA	Endometrium 2	4.23		0.00	
Endometrium 3AX	Endometrium 3	1.64		0.15	
Uterus 141XO		2.92		0.90	
Uterus 23XU	Mammary 1	12.91		3.31	

0.00= Negative

In the analysis of matching samples, higher expression of Pro132 is detected in prostate samples showing a high degree of tissue specificity for prostate tissue. These results confirm the tissue specificity results obtained with normal pooled samples (Table 8).

Furthermore, we compared the level of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual. This comparison provides an indication of specificity for the cancer stage (e.g. higher levels of mRNA expression in the cancer sample compared to the normal adjacent). Table 9 shows overexpression of Pro132 in 43% of the prostate matching samples tested (6 out of total of 14 prostate matching samples).

Altogether, the high level of tissue specificity, plus the mRNA differential expression in the prostate matching samples tested are believed to make Pro132 a good marker for diagnosing, monitoring, staging, imaging and treating prostate cancer.

**SEQ ID NO: 5 (Pro133)**

Table 10 shows the absolute numbers are relative levels of expression of Pro133 in 24  
 5 normal different tissues. All the values are compared to normal endometrium  
 (calibrator). These RNA samples are commercially pools, originated by pooling samples  
 of a particular tissue from different individuals.

**Table 10: Pro133**

Tissue	NORMAL
Adrenal Gland	0.01
Bladder	0.00
Brain	0.01
Cervix	0.11
Colon	0.09
Endometrium	1.00
Esophagus	0.03
Heart	0.00
Kidney	0.01
Liver	0.00
Lung	0.00
Mammary Gland	0.00
Muscle	0.00
Ovary	0.00
Pancreas	0.00
Prostate	112.99
Rectum	21.33
Small Intestine	0.00
Spleen	0.00
Stomach	0.00
Testis	0.03
Thymus	0.33
Trachea	0.13
Uterus	0.00

10 0=negative

The relative levels of expression in Table 10 show that Pro133 mRNA expression is  
 high in prostate compared with all other normal tissues analyzed.

The absolute numbers in Table 10 were obtained analyzing pools of samples of a  
 15 particular tissue from different individuals. They can not be compared to the absolute  
 numbers originated from RNA obtained from tissue samples of a single individual in  
 Table 11.

Table 11 shows the absolute numbers are relative levels of expression of Pro133 in matching samples and some unmatched samples. All the values are compared to normal endometrium (calibrator). A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same

5 tissue from the same individual.

Table 11: Pro133

Sample ID	Tissue	CANCER	PROSTATITIS & BENIGN HYPERPLASIA	MATCHING NORMAL ADJACENT	NORMAL
Pro53P	Prostate 1				51.45
ProC153	Prostate 2				0.90
Pro73P	Prostate 3				9.99
Pro77P	Prostate 4				6.17
Pro101XB	Prostate 5	69.59		79.07	
Pro65XB	Prostate 6	20.68		40.79	
Pro23B	Prostate 7	80.73		56.30	
Pro90XB	Prostate 8	32.33		31.89	
Pro91XB	Prostate 9	58.28		16.51	
Pro125XB	Prostate 10	3.42		5.12	
Pro12B	Prostate 11	27.57		1.08	
Pro69XB	Prostate 12	8.51		3.81	
Pro109XB	Prostate 13	7.89		10.45	
Pro13XB	Prostate 14	1.39		3.53	
Pro34B	Prostate 15	95.34		59.71	
Pro78XB	Prostate 16	45.73		47.01	
Pro84XB	Prostate 17	100.43		17.27	
Pro110	Prostate 18	5.48		40.79	
Pro326	Prostate 19	100.43		50.56	
Pro10R	Prostate 20 (prostatitis)		15.35		
Pro20R	Prostate 21 (prostatitis)		21.26		
Pro10P	Prostate 22 (BPH)		35.75		
Pro13P	Prostate 23 (BPH)		1.47		
Pro34P	Prostate 24 (BPH)		2.15		
Pro277P	Prostate 25 (BPH)		36.76		
Pro267A	Prostate 26 (BPH)		8.46		
Pro271A	Prostate 27 (BPH)		7.06		

Pro258	Prostate 28 (BPH)		10.70		
Pro263C	Prostate 29 (BPH)		46.53		
Pro460Z	Prostate 30 (BPH)		30.70		
Pro65P	Prostate 31 (BPH)		10.09		
Pro705P	Prostate 32 (BPH)		2.60		
Pro784P	Prostate 33 (BPH)		13.83		
Pro855P	Prostate 34 (BPH)		50.04		
ProC003P	Prostate 35 (BPH)		2.67		
ProC032	Prostate 36 (BPH)		4.81		
ProC034P	Prostate 37 (BPH)		5.96		
Kidney 716K	Kidney 1	1.18		0.01	
Kidney 106XD	Kidney 2	0.00		0.00	
Kidney 109XD	Kidney 3	1.59		0.00	
Kidney 107XD	Kidney 4	0.00		0.00	
Bladder 32XK	Bladder 1	0.03		0.04	
Bladder 46XK	Bladder 2	0.00		0.04	
Bladder 66X	Bladder 3	0.04		0.15	
Lung LC80	Lung 1	0.00		0.00	
Lung 143L	Lung 2	0.00		0.00	
Lung 205L	Lung 3	0.00		0.00	
Stomach 88S	Stomach 1	0.09		0.00	
Stomach 115S	Stomach 2	0.92		0.00	
Stomach 15S	Stomach 3	0.00		0.00	
Colon DC22	Colon 1	1.49		0.38	
Colon AC119	Colon 2	1.51		0.00	
Colon AS12	Colon 3	0.50		0.00	
Mammary 162X	Mammary 1	0.47		0.00	
Mammary 173M	Mammary 2	0.00		0.00	
Mammary 19DN	Mammary 3	0.12		0.00	
Ovary 1118	Ovary 1	0.00			
Ovary 32RA	Ovary 2			0.00	
Ovary G010	Ovary 3	0.03		0.00	
Ovary G021	Ovary 4	0.49		0.00	
Ovary10050	Ovary 5	0.00			
Ovary C057	Ovary 6	0.00			

Cervix KS83	Cervix 1	0.06		8.43	
Cervix NK23	Cervix 2	0.77		0.89	
Endometrium 28XA	Endometrium 1	0.49		0.07	
Endometrium 3AX	Endometrium 2	0.00		0.00	
Endometrium 10479	Endometrium 3	4.64		0.00	
Endometrium 12XA	Endometrium 4	4.59		0.12	
Endometrium 5XA	Endometrium 5	0.00		0.69	
Uterus 141XO	Uterus 1	0.57		0.00	
Uterus 135XO	Uterus 2	0.00		0.00	

0.00= Negative

In the analysis of matching samples, higher expression of Pro133 is detected in prostate samples showing a high degree of tissue specificity for prostate tissue. These results confirm the tissue specificity results obtained with normal pooled samples (Table 10).

Furthermore, we compared the level of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual. This comparison provides an indication of specificity for the cancer stage (e.g. higher levels of mRNA expression in the cancer sample compared to the normal adjacent). Table 11 shows overexpression of Pro133 in 47% of the prostate matching samples tested (7 out of total of 15 prostate matching samples).

Altogether, the high level of tissue specificity, plus the mRNA differential expression in the prostate matching samples tested are believed to make Pro133 a good marker for diagnosing, monitoring, staging, imaging and treating prostate cancer.

#### SEQ ID NO:10 (Pro135)

Table 12 shows the absolute numbers which are relative levels of expression of Pro135 in 24 normal different tissues. All the values are compared to normal endometrium (calibrator). These RNA samples are commercially pools, originated by pooling samples of a particular tissue from different individuals.

**Table 12: Pro135**

Tissue	NORMAL
Adrenal Gland	0.06

137

Bladder	0.00
Brain	0.21
Cervix	0.02
Colon	0.02
Endometrium	1.00
Esophagus	0.00
Heart	0.05
Kidney	0.02
Liver	0.02
Lung	0.51
Mammary Gland	0.18
Muscle	0.04
Ovary	0.40
Pancreas	0.70
Prostate	0.64
Rectum	0.33
Small Intestine	0.04
Spleen	1.00
Stomach	0.08
Testis	0.42
Thymus	1.07
Trachea	0.54
Uterus	0.99

0=negative

The relative levels of expression in Table 12 show that Pro135 mRNA expression is high in prostate compared with all other normal tissues analyzed.

5 The absolute numbers in Table 12 were obtained analyzing pools of samples of a particular tissue from different individuals. They can not be compared to the absolute numbers originated from RNA obtained from tissue samples of a single individual in Table 13.

10 Table 13 shows the absolute numbers which are relative levels of expression of Pro135 in matching samples and some unmatched samples. All the values are compared to normal endometrium (calibrator). A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

Table 13: Pro135

15

Sample ID	Tissue	CANCER	PROSTATITIS & BENIGN HYPERPLASIA	MATCHING NORMAL ADJACENT	NORMAL
Pro77P	Prostate 1				0.08
Pro101XB	Prostate 2	1.08		0.88	

Pro65XB	Prostate 3	0.97		0.52	
Pro23B	Prostate 4	2.00		0.58	
Pro90XB	Prostate 5	0.27		0.41	
Pro91XB	Prostate 6	0.56		0.43	
Pro125XB	Prostate 7	0.20		0.70	
Pro12B	Prostate 8	0.73		0.00	
Pro69XB	Prostate 9	0.41		0.34	
Pro10P	Prostate 10 (BPH)		0.24		
Pro13P	Prostate 11 (BPH)		0.18		
Pro34P	Prostate 12 (BPH)		0.08		
Pro277P	Prostate 13 (BPH)		0.41		
Pro267A	Prostate 14 (BPH)		0.16		
Pro271A	Prostate 15 (BPH)		0.00		
Testis 647T	Testis 1	0.28		0.48	
Bladder 46XK	Bladder 1	0.05		0.06	
Bladder TR14	Bladder 2	0.93		0.42	
Bladder 46K	Bladder 3				0.30
Liver 15XA	Liver 1	0.23		0.12	
Colon SG33	Colon 1	0.59		0.25	
Mammary 12B	Mammary 1	0.26			
Mammary A04	Mammary 2				0.17
Ovary 32RA	Ovary 1				1.49
Ovary 1461	Ovary 2	0.24			
Uterus 135XO	Uterus 1	1.09		2.26	

0.00= Negative

In the analysis of matching samples, higher expression of Pro135 is detected in prostate samples showing a high degree of tissue specificity for prostate tissue. These results confirm the tissue specificity results obtained with normal pooled samples (Table 12).

Furthermore, we compared the level of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual. This comparison provides an indication of specificity for the cancer stage (e.g. higher levels of mRNA expression in

the cancer sample compared to the normal adjacent). Table 13 shows overexpression of Pro135 in 38% of the prostate matching samples tested (3 out of total of 8 prostate matching samples).

- 5 Altogether, the high level of tissue specificity, plus the mRNA differential expression in the prostate matching samples tested are believed to make Pro135 a good marker for diagnosing, monitoring, staging, imaging and treating prostate cancer.

#### SEQ ID NO:13 (Pro158)

- 10 Table 13 shows the absolute numbers which are relative levels of expression of Pro158 in 24 normal different tissues. All the values are compared to normal endometrium (calibrator). These RNA samples are commercially pools, originated by pooling samples of a particular tissue from different individuals.

15

**Table 14: Pro158**

Tissue	NORMAL
Adrenal Gland	0.24
Bladder	0.17
Brain	1.64
Cervix	0.61
Colon	0.18
Endometrium	1.00
Esophagus	0.08
Heart	0.14
Kidney	0.51
Liver	0.28
Lung	2.79
Mammary Gland	0.44
Muscle	0.06
Ovary	1.79
Pancreas	0.33
Prostate	112.99
Rectum	0.72
Small Intestine	1.47
Spleen	2.51
Stomach	0.83
Testis	7.14
Thymus	1.25
Trachea	0.38



140

Uterus	2.77
--------	------

0=negative

The relative levels of expression in Table 14 show that Pro158 mRNA expression is high in prostate compared with all other normal tissues analyzed.

5

**SEQ ID NO:14 (Pro131)**

Table 15 shows the absolute numbers which are relative levels of expression of Pro131 in 24 normal different tissues. All the values are compared to normal testis (calibrator). These RNA samples are commercially pools, originated by pooling samples of a particular tissue from different individuals.

10

**Table 15: Pro131**

Tissue	NORMAL
Adrenal Gland	0.00
Bladder	0.00
Brain	0.00
Cervix	0.00
Colon	0.01
Endometrium	0.02
Esophagus	0.01
Heart	0.00
Kidney	0.00
Liver	0.00
Lung	0.00
Mammary Gland	0.01
Muscle	0.01
Ovary	0.00
Pancreas	0.00
Prostate	39.53
Rectum	0.00
Small Intestine	0.00
Spleen	0.01
Stomach	0.14
Testis	1.00
Thymus	0.02
Trachea	0.03
Uterus	0.00

0=negative

15

The relative levels of expression in Table 15 show that Pro131 mRNA expression is high in prostate compared with all other normal tissues analyzed.

The absolute numbers in Table 15 were obtained analyzing pools of samples of a particular tissue from different individuals. They can not be compared to the absolute numbers originated from RNA obtained from tissue samples of a single individual in Table 16.

- 5 Table 16 shows the absolute numbers which are relative levels of expression of Pro131 in matching samples and some unmatched samples. All the values are compared to normal testis (calibrator). A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

10

**Table 16: Pro131**

Sample ID	Tissue	CANCER	PROSTATITIS & BENIGN HYPERPLASIA	MATCHING NORMAL ADJACENT	NORMAL
Pro53P	Prostate 1				10.37
Pro73P	Prostate 2				1.28
Pro77P	Prostate 3				3.28
Pro101XB	Prostate 4	31.67		23.18	
Pro65XB	Prostate 5	5.01		9.92	
Pro23B	Prostate 6	14.47		10.67	
Pro90XB	Prostate 7	2.66		10.48	
Pro91XB	Prostate 8	15.14		3.11	
Pro125XB	Prostate 9	4.30		1.60	
Pro12B	Prostate 10	9.71		0.41	
Pro69XB	Prostate 11	1.44		0.28	
Pro109XB	Prostate 12	0.52		1.62	
Pro110	Prostate 13	0.11		2.23	
Pro13XB	Prostate 14	0.32		1.32	
Pro326	Prostate 15	1.06		0.80	
Pro34B	Prostate 16	17.27		6.89	
Pro78XB	Prostate 17	9.82		2.11	
Pro84XB	Prostate 18	17.21		0.87	
Pro10R	Prostate 19 (prostatitis)		0.70		
Pro20R	Prostate 20 (prostatitis)		0.66		
Pro10P	Prostate 21 (BPH)		3.89		
Pro13P	Prostate 22 (BPH)		0.48		
Pro34P	Prostate 23 (BPH)		2.01		
Pro277P	Prostate 24 (BPH)		11.92		
Pro267A	Prostate 25 (BPH)		0.96		

Pro271A	Prostate 26 (BPH)		0.94		
Pro258	Prostate 27 (BPH)		0.19		
Pro263C	Prostate 28 (BPH)		4.00		
Pro34P	Prostate 29 (BPH)		1.23		
Pro460Z	Prostate 30 (BPH)		1.32		
Pro65P	Prostate 31 (BPH)		0.65		
Pro705P	Prostate 32 (BPH)		0.00		
Pro784P	Prostate 33 (BPH)		0.83		
Pro83P	Prostate 34 (BPH)		0.28		
Pro855P	Prostate 35 (BPH)		2.85		
ProC003P	Prostate 36 (BPH)		0.03		
ProC032	Prostate 37 (BPH)		0.47		
Bladder 46K	Bladder 1	0.00		0.00	
Bladder 46XK	Bladder 2	0.04		0.02	
Bladder TR14	Bladder 3	0.00		0.00	
Bladder 32XK	Bladder 4	0.01		0.00	
Bladder 66X	Bladder 5	0.01		0.00	
Kidney 106XD	Kidney 1	0.00		0.00	
Kidney 107XD	Kidney 2	0.00		0.00	
Kidney 109XD	Kidney 3	0.00		0.00	
Kidney 10XD	Kidney 4	0.00		0.00	
Testis 647T	Testis 1	0.00		0.09	
Testis 39X	Testis 2	0.00		0.03	
Testis 663T	Testis 3	0.00		0.03	
Liver 15XA	Liver 1	0.00		0.00	
Lung 143L	Lung 1	0.00		0.00	
Lung 205L	Lung 2	0.00		0.04	
Lung 223L	Lung 3	0.00		0.00	
Pancreas 77X	Pancreas 1	0.00		0.01	
Colon SG33	Colon 1	0.00		0.00	
Colon 132C	Colon 2	0.01		0.00	
Colon AC19	Colon 3	0.00		0.00	
Colon AS12	Colon 4	0.00		0.00	
Stomach 115S	Stomach 1	0.00		0.00	
Stomach 15S	Stomach 2	0.00		0.00	

Mammary 12B	Mammary 1	0.01			
Mammary A04	Mammary 2			0.00	
Mammary 162X	Mammary 3	0.05		0.00	
Mammary 173M	Mammary 4	0.00		0.02	
Ovary 32RA	Ovary 1			0.00	
Ovary 1461	Ovary 2	0.00			
Ovary 103X	Ovary 3	0.00		0.00	
Ovary A084	Ovary 4	0.00		0.00	
Endometrium 12XA	Endometrium 1	0.00		0.00	
Endometrium 28XA	Endometrium 2	0.01		0.01	
Uterus 135XO	Uterus 1	0.00		0.00	
Uterus 135XO	Uterus 2	0.00		0.00	
Uterus 141X	Uterus 3	0.00		0.00	

0.00= Negative

In the analysis of matching samples, higher expression of Pro131 is detected in prostate samples showing a high degree of tissue specificity for prostate tissue. These results confirm the tissue specificity results obtained with normal pooled samples (Table 15).

Furthermore, we compared the level of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual. This comparison provides an indication of specificity for the cancer stage (e.g. higher levels of mRNA expression in the cancer sample compared to the normal adjacent). Table 16 shows overexpression of Pro131 in 47% of the prostate matching samples tested (7 out of total of 15 prostate matching samples).

Altogether, the high level of tissue specificity, plus the mRNA differential expression in the prostate matching samples tested are believed to make Pro131 a good marker for diagnosing, monitoring, staging, imaging and treating prostate cancer.

### Example 3: Protein Expression

The PSNA is amplified by polymerase chain reaction (PCR) and the amplified DNA fragment encoding the PSNA is subcloned in pET-21d for expression in E. coli. In addition to the PSNA coding sequence, codons for two amino acids, Met-Ala, flanking the NH<sub>2</sub>-terminus of the coding sequence of PSNA, and six histidines, flanking the COOH-terminus of the coding sequence of PSNA, are incorporated to serve as initiating Met/restriction site and purification tag, respectively.

An over-expressed protein band of the appropriate molecular weight may be observed on a Coomassie blue stained polyacrylamide gel. This protein band is confirmed by Western blot analysis using monoclonal antibody against 6X Histidine tag.

5 Large-scale purification of PSP was achieved using cell paste generated from 6-liter bacterial cultures, and purified using immobilized metal affinity chromatography (IMAC). Soluble fractions that had been separated from total cell lysate were incubated with a nickel chelating resin. The column was packed and washed with five column volumes of wash buffer. PSP was eluted stepwise with various concentration imidazole buffers.

#### 10 **Example 4: Protein Fusions**

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector. For example, if pC4  
15 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 2, is ligated into this BamHI site. Note that the polynucleotide is cloned without  
20 a stop codon, otherwise a fusion protein will not be produced. If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. *See, e. g.,* WO 96/34891.

#### 25 **Example 5: Production of an Antibody from a Polypeptide**

In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine  
30 serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 g/ml of streptomycin. The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any

suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP20), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands *et al.*,

5 *Gastroenterology* 80: 225-232 (1981).

The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide.

Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes  
10 use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability  
15 to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies. Using the Jameson-Wolf methods the following epitopes were predicted. (Jameson and Wolf, CABIOS, 4(1), 181-186, 1988, the contents of which are incorporated by  
20 reference).

DEX84_4.aa		Antigenicity Index (Jameson-Wolf)	
(SEQ ID NO:24)			
25	positions	AI avg	length
	115-142	1.09	28
	101-111	0.95	11
DEX84_12.aa		Antigenicity Index (Jameson-Wolf)	
(SEQ ID NO:26)			
30	positions	AI avg	length
	225-236	1.34	12
	68-78	1.06	11
35	306-344	1.03	39
	276-295	1.02	20
	153-164	1.00	12
	201-214	0.95	14
DEX84_13.aa		Antigenicity Index (Jameson-Wolf)	
(SEQ ID NO:27)			
40	positions	AI avg	length
	238-267	1.32	30
	93-112	1.05	20
	52-77	0.96	26

	382-394	0.95	13
	5-19	0.93	15
	DEX84_16.aa Antigenicity Index(Jameson-Wolf)		
	(SEQ ID NO:29)		
5	positions	AI avg	length
	261-272	1.47	12
	222-245	0.93	24
	DEX84_17.aa Antigenicity Index(Jameson-Wolf)		
	(SEQ ID NO:30)		
10	positions	AI avg	length
	10-38	1.00	29
	DEX84_19.aa Antigenicity Index(Jameson-Wolf)		
	(SEQ ID NO:31)		
	positions	AI avg	length
15	103-125	1.19	23
	27-43	0.92	17

#### Example 6: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide

20 RNA is isolated from individual patients or from a family of individuals that have a phenotype of interest. cDNA is then generated from these RNA samples using protocols known in the art. *See*, Sambrook (2001), *supra*. The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO:1-78. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120  
25 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky *et al.*, *Science* 252(5006): 706-9 (1991). *See also* Sidransky *et al.*, *Science* 278(5340): 1054-9 (1997).

PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies).  
30 The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations are then cloned and sequenced to validate the results of the direct sequencing. PCR products is cloned into T-tailed vectors as described in Holton *et al.*, *Nucleic Acids Res.*, 19: 1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected  
35 individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements may also be determined. Genomic clones are nick-translated with digoxigenin deoxyuridine 5' triphosphate (Boehringer Mannheim), and FISH is performed as described in Johnson *et al.*, *Methods Cell Biol.* 35: 73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human  
40 cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C-and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. *Id.* Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

**Example 7: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample**

Antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 ug/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described above. The wells are blocked so that non-specific binding of the polypeptide to the well is reduced. The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbound polypeptide. Next, 50 µl of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbound conjugate. 75 µl of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution are added to each well and incubated 1 hour at room temperature.

The reaction is measured by a microtiter plate reader. A standard curve is prepared, using serial dilutions of a control sample, and polypeptide concentrations are plotted on the X-axis (log scale) and fluorescence or absorbance on the Y-axis (linear scale). The concentration of the polypeptide in the sample is calculated using the standard curve.

**Example 8: Formulating a Polypeptide**



The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1  $\mu\text{g/kg/day}$  to 10  $\text{mg/kg/day}$  of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01  $\text{mg/kg/day}$ , and most preferably for humans between about 0.01 and 1  $\text{mg/kg/day}$  for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1  $\mu\text{g/kg/hour}$  to about 50  $\text{mg/kg/hour}$ , either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracisternally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semipermeable polymer matrices in the form of shaped articles, e. g., films, or microcapsules. Sustained-release matrices include polylactides (U. S. Pat. No.3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22: 547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15: 167-277 (1981), and R. Langer, Chem. Tech. 12: 98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-

release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE Epstein et al., Proc. Natl. Acad. Sci. USA 82: 3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77: 4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 5 142,641; Japanese Pat. Appl. 83-118008; U. S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

10 For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, I. e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation.

15 For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides. Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more 20 preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that 25 enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e. g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic 30 polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrans; chelating agents such

as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e. g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper peaceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1 % (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container (s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

#### **Example 9: Method of Treating Decreased Levels of the Polypeptide**

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 µg/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided above.

5

**Example 10: Method of Treating Increased Levels of the Polypeptide**

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

10 For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided above.

15

**Example 11: Method of Treatment Using Gene Therapy**

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and  
20 separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e. g., Ham's F12 media, with 10% FBS, penicillin and  
25 streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks. pMV-7 (Kirschmeier, P. T. et al.,  
30 DNA, 7: 219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB 101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+aml2 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media.

If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

#### **Example 12: Method of Treatment Using Gene Therapy-In Vivo**

Another aspect of the present invention is using in vivo gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the

introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences into an animal to increase or decrease the expression of the polypeptide.

The polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known in the art, see, for example, WO 90/11092, WO 98/11779; U. S. Patent No. 5,693,622; 5,705,151; 5,580,859; Tabata H. et al. (1997) *Cardiovasc. Res.* 35 (3): 470-479, Chao J et al. (1997) *Pharmacol. Res.* 35 (6): 517-522, Wolff J. A. (1997) *Neuromuscul. Disord.* 7 (5): 314-318, Schwartz B. et al. (1996) *Gene Ther.* 3 (5): 405-411, Tsurumi Y. et al. (1996) *Circulation* 94 (12): 3281-3290 (incorporated herein by reference).

The polynucleotide constructs may be delivered by any method that delivers injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, prostate, liver, intestine and the like). The polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides of the present invention may also be delivered in liposome formulations (such as those taught in Felgner P. L. et al. (1995) *Ann. NY Acad. Sci.* 772: 126-139 and Abdallah B. et al. (1995) *Biol. Cell* 85 (1): 1-7) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

The polynucleotide construct can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, prostate, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and

connective tissue. Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly  
5 the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved  
10 in non-differentiated or less completely differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. In vivo muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05  $\mu\text{g/kg}$  body weight to about 50  $\text{mg/kg}$  body  
15 weight. Preferably the dosage will be from about 0.005  $\text{mg/kg}$  to about 20  $\text{mg/kg}$  and more preferably from about 0.05  $\text{mg/kg}$  to about 5  $\text{mg/kg}$ . Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be determined by those of ordinary skill in the art and may depend on the condition being  
20 treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to prostates or bronchial tissues, throat or mucous membranes of the nose. In addition, naked polynucleotide constructs can be delivered to arteries during  
25 angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle in vivo is determined as follows. Suitable template DNA for production of mRNA coding for polypeptide of the present invention is prepared in accordance with a standard recombinant DNA methodology. The template DNA, which may be either circular or  
30 linear, is either used as naked DNA or complexed with liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the

anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

After an appropriate incubation time (e. g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 um cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A time course for protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice.

The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA.

15

### Example 13: Transgenic Animals

The polypeptides of the invention can also be expressed in transgenic animals. Animals of any species, including, but not limited to, mice, rats, rabbits, hamsters, guinea pigs, pigs, micro-pigs, goats, sheep, cows and non-human primates, e. g., baboons, monkeys, and chimpanzees may be used to generate transgenic animals. In a specific embodiment, techniques described herein or otherwise known in the art, are used to express polypeptides of the invention in humans, as part of a gene therapy protocol.

Any technique known in the art may be used to introduce the transgene (i. e., polynucleotides of the invention) into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection (Paterson et al., Appl. Microbiol. Biotechnol. 40: 691-698 (1994); Carver et al., Biotechnology (NY) 11: 1263-1270 (1993); Wright et al., Biotechnology (NY) 9: 830-834 (1991); and Hoppe et al., U. S. Pat. No. 4,873,191 (1989)); retrovirus mediated gene transfer into germ lines (Van der Putten et al., Proc. Natl. Acad. Sci., USA 82: 6148-6152 (1985)), blastocysts or embryos; gene targeting in embryonic stem cells (Thompson et al., Cell 56: 313-321 (1989)); electroporation of cells or embryos (Lo, 1983, Mol Cell. Biol. 3: 1803-1814 (1983)); introduction of the polynucleotides of the invention using a gene gun (see, e. g., Ulmer et al., Science 259: 1745 (1993); introducing nucleic acid



constructs into embryonic pluripotent stem cells and transferring the stem cells back into the blastocyst; and sperm mediated gene transfer (Lavitrano et al., Cell 57: 717-723 (1989); etc. For a review of such techniques, see Gordon, "Transgenic Animals," Intl. Rev. Cytol. 115: 171-229 (1989), which is incorporated by reference herein in its entirety.

Any technique known in the art may be used to produce transgenic clones containing polynucleotides of the invention, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (Campell et al., Nature 380: 64-66 (1996); Wilmut et al., Nature 385: 810813 (1997)).

The present invention provides for transgenic animals that carry the transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, i.e., mosaic animals or chimeric. The transgene may be integrated as a single transgene or as multiple copies such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching of Lasko et al. (Lasko et al., Proc. Natl. Acad. Sci. USA 89: 6232-6236 (1992)). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. When it is desired that the polynucleotide transgene be integrated into the chromosomal site of the endogenous gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type, by following, for example, the teaching of Gu et al. (Gu et al., Science 265: 103-106 (1994)). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using

techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and reverse transcriptase-PCR (rt-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

#### **Example 14: Knock-Out Animals**

Endogenous gene expression can also be reduced by inactivating or "knocking out" the gene and/or its promoter using targeted homologous recombination. (E. g., see Smithies et al., Nature 317: 230-234 (1985); Thomas & Capecchi, Cell 51: 503-512 (1987); Thompson et al., Cell 5: 313-321 (1989); each of which is incorporated by reference herein in its entirety). For example, a mutant, non-functional polynucleotide of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous polynucleotide sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention in vivo. In another embodiment, techniques known in the art are used to generate knockouts in cells that

contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive  
5 targeted gene (e. g., see Thomas & Capecchi 1987 and Thompson 1989, *supra*). However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site in vivo using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to  
10 express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (e. g., knockouts) are administered to a patient in vivo. Such cells may be obtained from the patient (i. e., animal, including human) or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (e. g., lymphocytes), adipocytes, muscle cells, endothelial  
15 cells etc. The cells are genetically engineered in vitro using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, e. g., by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or  
20 transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered  
25 cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, e. g., in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, e. g., genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or  
30 vascular graft. (See, for example, Anderson et al. U. S. Patent No. 5,399,349; and Mulligan & Wilson, U. S. Patent No. 5,460,959 each of which is incorporated by reference herein in its entirety).

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange  
5 of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated  
10 with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

All patents, patent publications, and other published references mentioned herein are hereby incorporated by reference in their entireties as if each had been individually and specifically incorporated by reference herein. While preferred illustrative  
15 embodiments of the present invention are described, one skilled in the art will appreciate that the present invention can be practiced by other than the described embodiments, which are presented for purposes of illustration only and not by way of limitation. The present invention is limited only by the claims that follow.

**What is claimed is:**

1. An isolated nucleic acid molecule selected from the group consisting of:
  - (a) a nucleic acid molecule comprising a nucleic acid sequence that encodes an amino acid sequence of SEQ ID NO:23-31;
  - 5 (b) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO:1-22;
  - (c) a nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of (a) or (b);
  - (d) a nucleic acid molecule that is substantially similar to the nucleic acid molecule of (a) or (b);
  - 10 (e) a nucleic acid molecule that is an allelic variant of (a) or (b);
  - (f) a nucleic acid molecule that is a part of any one of (a), (b), (c), (d) or (e); and
  - (g) a nucleic acid molecule of any one of (a), (b), (c), (d), (e) or (f) that is modified to include one or more nonnative internucleoside bonds, post-synthetic modifications and/or altered nucleotide analogues.
  - 15
2. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is a prostate specific nucleic acid.
- 20 3. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is a cDNA.
4. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is genomic DNA.
- 25 5. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is a mammalian nucleic acid molecule.
6. The nucleic acid molecule according to claim 5, wherein the nucleic acid molecule is a human nucleic acid molecule.
- 30

7. The nucleic acid molecule according to claim 1, wherein said selectively hybridizing nucleic acid molecule of (c) hybridizes to the nucleic acid of (a) or (b) under conditions of low stringency, moderate stringency or high stringency.

5           8. The nucleic acid molecule according to claim 7, wherein said selectively hybridizing nucleic acid molecule of (c) hybridizes to the nucleic acid molecule of (a) or (b) under conditions of high stringency.

9. The nucleic acid molecule according to claim 1, wherein said substantially  
10 similar nucleic acid molecule of (d) has at least 60% sequence identity to the nucleic acid molecule of (a) or (b).

10. The nucleic acid molecule according to claim 9, wherein said sequence  
identity is at least 80%.

15

11. The nucleic acid molecule according to claim 10, where said sequence  
identity is at least 90%.

12. The nucleic acid molecule according to claim 1, wherein said substantially  
20 similar nucleic acid molecule of (d) has at least 60% sequence identity over at least 100 nucleotides to the nucleic acid molecule of (a) or (b).

13. The nucleic acid molecule according to claim 12, wherein said substantially  
similar nucleic acid molecule of (d) has at least 60% sequence identity to the entire  
25 length of the nucleic acid molecule of (a) or (b).

14. The nucleic acid molecule according to claim 1, wherein said substantially  
similar nucleic acid molecule of (d) is derived from a mammalian species.

30           15. The nucleic acid molecule according to claim 14, wherein said substantially  
similar nucleic acid molecule of (d) is derived from a primate species.

16. The nucleic acid molecule according to claim 1, wherein said allelic variant of (e) derived from a human.

17. The nucleic acid molecule according to claim 16, wherein said allelic variant  
5 of (e) is a single nucleotide polymorphism (SNP).

18. The nucleic acid molecule according to claim 1, wherein said part comprises at least 17 nucleotides.

10 19. The nucleic acid molecule according to claim 18, wherein said part comprises a part of the nucleotide sequence selected from the group consisting of SEQ ID NO:1-22.

20. The nucleic acid molecule according to claim 18, wherein said part comprises a part of the nucleotide sequence encoding the amino acid sequence selected from the  
15 group consisting of SEQ ID NO:23-31.

21. The nucleic acid molecule according to claim 18, wherein said part is detectably labeled.

20 22. The nucleic acid molecule according to claim 1, wherein said modified nucleic acid molecule comprises a radiolabeled nucleotide analogue, a fluorescent nucleotide analogue or a hapten-conjugated nucleotide.

23. The nucleic acid molecule according to claim 1, wherein said modified  
25 nucleic acid molecule is a peptide nucleic acid.

24. The nucleic acid molecule according to claim 1, further comprising a nucleic acid sequence encoding a heterologous polypeptide, optionally comprising one or more expression control sequences operably linked to the nucleic acid molecule.

30

25. The nucleic acid molecule according to claim 1, further comprising one or more expression control sequences operably linked to the nucleic acid molecule.

26. A vector comprising the nucleic acid molecule according to either of claims 1 or 25.

27. The vector according to claim 26, wherein said vector is a replicable vector.

28. The vector according to claim 26, wherein said vector is a plasmid vector.

29. The vector according to claim 26, wherein said vector is a viral vector.

30. A host cell comprising the vector according to claim 26.

31. A host cell comprising the nucleic acid molecule according to either of claims 1 or 25.

32. The host cell according to claim 31, wherein the host cell is selected from the group consisting of a bacterial cell, an insect cell, a yeast cell, a plant cell and a mammalian cell.

33. The host cell according to claim 32, wherein the host cell is a human cell.

34. A method for producing a polypeptide encoded by the nucleic acid molecule according to claim 1, comprising the steps of

- (a) providing a host cell comprising the nucleic acid molecule operably linked to one or more expression control sequences, and
- (b) incubating the host cell under conditions in which the polypeptide is produced.

35. The method according to claim 34, further comprising the step of isolating the polypeptide.

36. The polypeptide encoded by the nucleic acid molecule according to claim 1.



37. An isolated polypeptide selected from the group consisting of:
- (a) a polypeptide comprising an amino acid sequence of SEQ ID NO:23-31;
  - (b) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO:1-22;
  - 5 (c) a polypeptide mutant (mutein) of (a) or (b), wherein said mutein comprises at least one amino acid insertion, deletion or substitution compared to the amino acid sequence of (a) or (b) and wherein said mutein exhibits at least 60% sequence identity to the amino acid sequence of (a) or (b);
  - 10 (d) a polypeptide which is homologous to (a) or (b), which exhibits at least 60% sequence identity to the amino acid sequence of (a) or (b);
  - (e) a polypeptide that is an allelic variant of (a) or (b);
  - (f) a polypeptide that is a fragment of (a), (b), (c), (d) or (e);
  - (g) a polypeptide that is a derivative of any one of (a), (b), (c), (d), (e) or (f);
  - 15 and
  - (h) a polypeptide that is an analog of any one of (a), (b), (c), (d), (e) or (f).
38. The polypeptide according to claim 37, wherein said polypeptide is a prostate specific polypeptide (PSP).
- 20 39. The polypeptide according to claim 38, wherein said polypeptide is a mammalian polypeptide.
40. The polypeptide according to claim 39, wherein said polypeptide is a human
- 25 polypeptide.
41. The polypeptide according to claim 37, wherein said mutein or said homologous polypeptide exhibits at least 80% sequence identity to the polypeptide of (a) or (b).
- 30 42. The polypeptide according to claim 37, wherein said allelic variant is encoded by a nucleic acid molecule comprising an SNP.

43. The polypeptide according to claim 37, wherein said fragment comprises at least six amino acids.

44. The polypeptide according to claim 43, wherein said fragment comprises a  
5 fragment of the amino acid sequence selected from the group consisting of SEQ ID NO:23-31.

45. The polypeptide according to claim 37, wherein said derivative has been  
10 labeled with a fluorophore, a radioactive isotope, a chemiluminescent agent, an enzyme or with an antiligand.

46. The polypeptide according to claim 37, wherein said analog comprises a non-  
natural amino acid or a non-native inter-residue bond compared to the naturally occurring  
15 polypeptide.

47. The polypeptide according to claim 37, wherein said polypeptide further  
comprises a heterologous amino acid sequence.

48. The polypeptide according to claim 47, wherein said heterologous amino acid  
20 sequence facilitates purification or visualization of the polypeptide.

49. An antibody or fragment thereof that specifically binds to a polypeptide  
according to claim 37.

25 50. The antibody according to claim 49, which is a monoclonal antibody or binding fragment thereof.

51. The antibody according to claim 50, which is detectably labeled.

30 52. The antibody according to claim 50, which is a primatized antibody, a humanized antibody or a human antibody.

53. The antibody according to claim 50, which is conjugated to a toxin.

54. A composition comprising the antibody according to claim 49 and a pharmaceutically acceptable carrier.

5 55. A composition comprising the polypeptide according to claim 37 and a pharmaceutically acceptable carrier.

56. The composition according to claim 55, further comprising a protein stabilizer.

10

57. The composition according to claim 57, wherein said stabilizer is polyethylene glycol (PEG) or albumin.

58. A composition comprising the nucleic acid molecule according to claim 1  
15 and a pharmaceutically acceptable carrier.

59. A method for determining the presence of a PSNA in a sample, comprising the steps of:

- 20 (a) contacting the sample with the nucleic acid molecule according to claim 1 under conditions in which the nucleic acid molecule will selectively hybridize to the PSNA; and
- (b) detecting hybridization of the nucleic acid molecule to a PSNA in the sample, wherein the detection of the hybridization indicates the presence of a PSNA in the sample.

25

60. The method according to claim 59, further comprising the step of quantifying the amount of the PSNA in the sample by determining the amount of hybridization of the nucleic acid molecule to the PSNA.

30 61. The method according to claim 59, wherein said sample is a sample suspected of comprising either a prostate cell or prostate tissue.

62. The method according to claim 59, wherein said sample is attached to a substrate.

63. The method according to claim 59, wherein said nucleic acid molecule is  
5 attached to a substrate.

64. The method according to claim 63, wherein said substrate forms a microarray or a macroarray.

10 65. A method for determining the presence of a PSNA in a sample, comprising the steps of:  
(a) contacting the sample with a first primer and a second primer derived from the nucleic acid molecule according to claim 1 to provide an amplification mixture, wherein said first and second primers are  
15 sufficient in length to selectively hybridize to a PSNA, and wherein said primers are capable of amplifying a detectable part of a PSNA if a PSNA is present in the sample;  
(b) subjecting the amplification mixture to nucleic acid amplification; and  
(c) detecting whether a part of a PSNA has been amplified, wherein  
20 detection of a part indicates the presence of a PSNA in the sample.

66. The method according to claim 65, wherein said nucleic acid amplification is polymerase chain reaction.

25 67. A method for determining the presence of a PSP in a sample, comprising the steps of:  
(a) contacting the sample with the antibody according to claim 49 under conditions in which the antibody will selectively bind to the PSP; and  
(b) detecting binding of the antibody to a PSP in the sample, wherein the  
30 detection of binding indicates the presence of a PSP in the sample.

68. The method according to claim 67, further comprising the step of quantifying the amount of the PSP in the sample by determining the amount of binding of the antibody to the PSP.

5           69. The method according to claim 68, wherein said sample is a sample suspected of comprising either a prostate cell or prostate tissue.

70. The method according to claim 67, wherein said sample is attached to a substrate.

10

71. The method according to claim 70, which is an ELISA.

72. A method for diagnosing the presence of prostate cancer in a patient, comprising the steps of:

15

- (a) providing a sample of a cell, a tissue or a body fluid from the patient;
- (b) determining the amount of prostate specific nucleic acid or prostate specific protein in the sample; and

20

- (c) comparing the amount of the nucleic acid molecule or the polypeptide in the sample to the amount of the nucleic acid molecule or the polypeptide in a normal control; wherein a difference in the amount of the nucleic acid molecule or the polypeptide in the sample compared to the amount of the nucleic acid molecule or the polypeptide in the normal control is associated with the presence of prostate cancer.

25

73. A method for diagnosing whether a prostate cancer has metastasized in a patient whose prostate cancer is not known to have metastasized, comprising the steps of

30

- (a) providing a sample of a cell, a tissue or a body fluid from the patient;
- (b) determining the amount of the prostate specific nucleic acid or prostate specific protein in the sample; and
- (c) comparing the amount of the nucleic acid molecule or the polypeptide in the sample to the amount of the nucleic acid molecule or the polypeptide in a control sample in which it is known that the prostate

cancer has not metastasized; wherein a difference between the amount of the nucleic acid molecule or the polypeptide in the sample compared to the amount of the nucleic acid molecule or the polypeptide in the control sample is associated with metastasis.

5

74. A method for staging prostate cancer in a patient, comprising the steps of
- (a) providing a sample of a cell, a tissue or a body fluid from the patient;
  - (b) determining the amount of the prostate specific nucleic acid or prostate specific protein in the sample; and
  - (c) comparing the amount of the nucleic acid molecule or the polypeptide in the sample to the amount of the nucleic acid molecule or the polypeptide in control samples, wherein the control samples are derived from at least two different stages of prostate cancer, which stages of prostate cancer for the control samples are known; wherein the amount of the nucleic acid molecule or the polypeptide in the sample that is most similar to the amount of the nucleic acid molecule or the polypeptide in one the control sample is indicative of the stage of the prostate cancer.

10

15

20

75. A method for monitoring a prostate cancer to determine the onset of metastasis in a patient whose prostate cancer is not known to have metastasized, comprising the steps of
- (a) providing a sample of a cell, a tissue or a body fluid from the patient;
  - (b) periodically determining the amount of the prostate specific nucleic acid or prostate specific protein in the sample; and
  - (c) comparing the periodically determined amounts of the nucleic acid molecule or the polypeptide in the sample to the amount of the nucleic acid molecule or the polypeptide in a control sample in which it is known that the prostate cancer has not metastasized; wherein a change between one or more of the periodically determined amounts of the nucleic acid molecule or the polypeptide in the sample compared to the amount of the nucleic acid molecule or the polypeptide in the control sample is associated with the onset of metastasis.

25

30

76. A method for monitoring prostate cancer in a patient, comprising the steps of
- (a) providing a sample of a cell, a tissue or a body fluid from the patient;
  - (b) periodically determining the amount of the prostate specific nucleic acid or prostate specific protein in the sample; and
  - (c) comparing the periodically determined amounts of the nucleic acid molecule or the polypeptide in the sample to the amount of the nucleic acid molecule or the polypeptide in a control sample in which it is known that the prostate cancer has not metastasized; wherein an increase in one or more of the periodically determined amounts of the nucleic acid molecule or the polypeptide in the sample compared to the amount of the nucleic acid molecule or the polypeptide in the control sample is associated with the onset of metastasis or a progression of the stage of the disease, and wherein a decrease in one or more of the periodically determined amounts of the nucleic acid molecule or the polypeptide in the sample compared to the amount of the nucleic acid molecule or the polypeptide in the control sample is associated with remission of the disease or regression of the stage of the disease.
77. A method for monitoring a patient at risk of developing prostate cancer to whether prostate cancer has developed, comprising the steps of
- (a) providing a sample of a cell, a tissue or a body fluid from the patient;
  - (b) periodically determining the amount of the prostate specific nucleic acid or prostate specific protein in the sample; and
  - (c) comparing the periodically determined amounts of the nucleic acid molecule or the polypeptide in the sample to the amount of the nucleic acid molecule or the polypeptide in a control sample; wherein a change between one or more of the periodically determined amounts of the nucleic acid molecule or the polypeptide in the sample compared to the amount of the nucleic acid molecule or the polypeptide in the control sample is associated with the onset of prostate cancer.

78. The method according to claim 72, wherein the amount of the nucleic acid molecule is determined by PCR or nucleic acid hybridization.

5           79. The method according to claim 72, wherein the amount of the polypeptide is determined by radioimmunoassay (RIA), ELISA, Western blot or two-dimensional polyacrylamide gel electrophoresis (2D-PAGE).

80. The method according to claim 72, wherein the amount of more than one  
10   nucleic acid molecule or more than one polypeptide is determined for the patient.

81. A method to identify a genetic lesion in a patient with prostate cancer or in a patient at risk of developing prostate cancer, comprising the steps of:

- 15           (a) providing a genomic sample of the patient;
- (b) determining the genomic nucleic acid sequence from the patient for the nucleic acid molecule according to claim 2;
- (c) comparing the genomic nucleic acid sequence of the patient to the corresponding genomic nucleic acid sequence from a normal control; and
- 20           (d) identifying the genetic lesions in the genomic nucleic acid molecule of the patient.

82. A method for diagnosing the presence of a noncancerous prostate disorder in a patient, comprising the steps of:

- 25           (a) providing a sample of a cell, a tissue or a body fluid from the patient;
- (b) determining the amount of the prostate specific nucleic acid or prostate specific protein in the sample; and
- (c) comparing the amount of the nucleic acid molecule or the polypeptide in the sample to the amount of the nucleic acid molecule or the
- 30           polypeptide in a normal control; wherein a difference in the amount of the nucleic acid molecule or the polypeptide in the sample compared to the amount of the nucleic acid molecule or the polypeptide in the



normal control is associated with the presence of a noncancerous prostate disorder.

- 5           83. A method for producing an engineered prostate cell, comprising the steps of
- (a) providing a pluripotent cell;
  - (b) introducing a nucleic acid molecule according to claim 2 into the cell;
  - and
  - (c) growing the cell under conditions in which the cell differentiates into a prostate cell.

10

84. A method of treating a patient with prostate cancer, comprising the step of administering a composition comprising a prostate specific nucleic acid or prostate specific protein to a patient in need thereof, wherein said administration induces an immune response against the prostate cancer cell expressing the nucleic acid molecule or
- 15 polypeptide.

85. A vaccine comprising the polypeptide of claim 38.

- 20           86. An antisense molecule that is complementary to the nucleic acid molecule according to claim 1.

87. A method for treating prostate cancer, comprising the step of administering the antisense molecule according to claim 86 to a patient in need thereof, wherein said administration causes downregulation of the nucleic acid molecule.

25

88. A method for treating prostate cancer, comprising the step of administering the antibody according to claim 49 to a patient in need thereof, wherein said administration causes a decrease in the polypeptide levels.

- 30           89. The method according to claim 88, wherein said antibody is conjugated to a therapeutic agent.

90. The method according to claim 89, wherein said therapeutic agent is a toxin.

91. A method for identifying an agent that binds to a polypeptide according to claim 38, comprising the steps of:

- 5                   (a) contacting the polypeptide with the agent; and  
                  (b) determining if the polypeptide binds to the agent.

92. The method according to claim 91, further comprising the step of determining whether an agent that binds to the polypeptide modulates the expression or activity of the polypeptide.

10

93. A method for identifying an agent that downregulates the expression of the nucleic acid molecule according to claim 2, comprising the steps of:

- (a) contacting a cell that expresses the nucleic acid molecule with the agent; and  
15                   (b) determining whether the cell downregulates the expression of the nucleic acid molecule.

94. A method for imaging prostate cancer in a patient, comprising the steps of:

- (a) administering the antibody according to claim 49 to a patient in need of imaging; and  
20                   (b) detecting the prostate cancer in the patient.

95. The method according to claim 94, wherein the antibody is labeled.

25           96. A computer readable means for storing the nucleic acid sequences of the nucleic acid molecules according to claim 1.

97. A computer readable means for storing the amino acid sequences of the polypeptides according to claim 37.

30

98. A kit for detecting the presence of a PSNA in a sample comprising: (a) one or more of nucleic acid probes: SEQ ID NO:1-22 or nucleic acid sequences fully

complementary thereto; (b) instructions for hybridization of the nucleic acid probe with the nucleic acid molecules in the sample.

99. The kit of 98 wherein said probe is detectably labeled.

5

100. A kit for detecting a risk of cancer comprising: (a) a means for determining the presence of PSNA in a sample, and (b) a means for determining the level of expression of said PSNA.

101. A kit for detecting PSP antibody comprising: (a) an epitope of PSP; (b) a solid support which supports said epitope; and (c) a binding molecule which binds to a constant region in an antibody.

102. The kit of claim 101 wherein said binding molecule is an anti-human antibody.

103. A kit for detecting the presence of cancer in an individual comprising:  
(a) a first antibody with a first antigen binding site that immunoreacts with a peptide that has a sequence of approximately 16 contiguous amino acids from the amino acid sequence of SEQ ID NO:23-31;  
(b) a second antibody that immunoreacts said first antibody at a site other than said first antigen binding site.

104. The kit of claim 103 wherein either said first antibody or said second antibody has an attached identifying label.

105. The kit of claim 103 wherein at least one of said first antibody and said second antibody is a monoclonal antibody.

106. The test kit of claim 103 wherein said first antibody is a polyclonal antibody with an attached identifying label and said second antibody is a monoclonal antibody adherent to a solid surface.

## 175

107. The test kit of claim 106 wherein said attached identifying label is biotin.

108. A test kit for detecting antibodies indicative of a prostate cancer comprising a first polypeptide of claim 37.

5

109. The test kit of claim 108, further comprising a reagent for detecting a complex comprising a polypeptide of claim 1 and at least one antibody specifically bound thereto.

10

110. The test kit of claim 109, wherein the reagent is an anti-antibody.

111. The test kit of claim 109, further comprising an additional antibody that is specific for a second polypeptide of claim 1 wherein said first and said second polypeptides are different.

15

112. The test kit of claim 109, wherein the reagent for detecting the complex is labeled.

113. The test kit of claim 109, wherein the reagent is an enzymatic reagent.

20

114. The test kit of claim 112, wherein the polypeptide is labeled. with a paramagnetic ion or with a radioisotope.

## SEQUENCE LISTING

<110> Sun, Yongming  
 Recipon, Herve  
 Cafferkey, Robert  
 diaDexus, Inc.  
 Ali, Shujath

<120> Compositions and Methods Relating to Prostate Specific  
 Genes

<130> DEX-0240

<140>

<141>

<150> 60/233,746

<151> 2000-09-19

<160> 40

<170> PatentIn Ver. 2.1

<210> 1

<211> 1303

<212> DNA

<213> Homo sapiens

<400> 1

```

tcacaattta tgttcttctg ccatggcttc agccgggtccc tctgtttggg gttcctcact 60
tcccaaaaca tctctccctt tttttttata taaatgtgcc atggcgatga aggtttgttc 120
gttctctcga ttttgatgca ggattctttg actggctctg cacattttgt tagtgtagat 180
ttctgcagaa gcactatcc aggtgagagg ttgaataagt ggaggaaaag gcacataagc 240
ccaataagaa tatttatgtg tagcaggtaa atcagtgtga gaggaactg gtgagacaga 300
aagtataagg aggagaatca ttaaataaaa cttattgtaa gtgagattct gaagaaggaa 360
gagaagaaca gtgttaagta tcctggcaaa ggagcattag ggttacagat agaattgaag 420
ttgctcgtca gctgattgaa gataaggaga ttggcctgga ttatccagggt aggctcaatg 480
taatcaggaa gggcctttta agtgagagag ggaggcagaa gaggaagtca gagcgatgtg 540
ctgtgaaatc tactaccgtt tgctgggttt gaaaatggag aaaaagagtg aggaactgag 600
aaacatggat ggccttgagg acgtggaaaa gggctactga aatgggacga catgaactca 660
aggaggctat ttatgaccat gtcatttgca acatgaagaa agcttatctg gagtgaagt 720
aatgagacc aacagagatg agagaccgg agaaatcctg gttacactgc ttgaatcctg 780
tcagtcctat actggagtcc tgtaataaca aaataatagt aataatccct ctgtttctta 840
tgtttatgcc aacttcaaca aaaagaaact tgactaagag acaatataag aatttaagt 900
gtaattaaga aagaactctc caccacgggg aatgtgaaag gtatatgagt cccttttcac 960
gatgcatgt catgtctttt aaataagcca tactttatgt tcaataaaaa gagaataagc 1020
aggattcgca agagaacaca atcccttttt aactgctggg aagatacttt tagtcattaa 1080
tgactggacg acaatttggg acacatatat ggatattggc cggtttgtga tgatgtgatt 1140

```

```

gggcctctaa gtgacaacat tgttcctgt atagagttag tggcaagtgc atttataaaa 1200
ttggccatca tggctgttaa atttatgagt ctagaagtgt gcctctcaaa caagtatttg 1260
agaggttatc atgaagaaaa acaaaattaa aattattctg ctg 1303

```

&lt;210&gt; 2

&lt;211&gt; 1723

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; unsure

&lt;222&gt; (1040)

&lt;223&gt; a, c, g, or t

&lt;220&gt;

&lt;221&gt; unsure

&lt;222&gt; (1043)

&lt;223&gt; a, c, g, or t

&lt;400&gt; 2

```

gcatgatata ccatgcctat ttagagtagt tacattattt tcaagtctta ttgtaagagc 60
catttattgc ctttggccta aatgactcaa tataatatct ctgaaacttt tttttgacaa 120
attttggggc atgatgatga gagaaggggg tttgaaactt tctaataaga gttaacttag 180
agccatttaa gaaaggaaaa aacacaaatt atcagaaaaa caaaagtaag atcaagtgc 240
aaagtctctg ggcaagatg atgagaataa agaatatatg tttgtgactc atgggtggctt 300
ttactttgtt cttgcatttc tgagtacggg ttaacattta aagaatctac attatagata 360
acattttatt gcaagtaaat gtatttcaaa atttgttatt ggttttgtat gagattattc 420
tcagcctact tcattatcaa gctatattat tttattaatg tagtttgatg atcttacagc 480
aaagctgaaa gctgtatctt caaaatatgt ctatttgact aaaaagaagt tattcaacag 540
gagttattat ctatgaaaaa aatacaacag gaatataaaa aacttgaaga ggataaaaag 600
atgttggaag aagtaatat aaatcttaaa aaacatatgg aaagtacaca ttggtgaaga 660
cacattgggtg aagtacaaaa atataaattg gatctagaag aaagggaat tcaggcaata 720
gaaaaattag tagaaatccc tttaaagggt agtttgtaaa atcaggtaag tttatttata 780
atttgctttc atttatttca ctgcaaatta tttttggat atatatatat attgtgcttc 840
ctctgcctgt cttacagcaa tttgccttgc agagttctag gaaaaagggt gcatgtggtt 900
ttactttcaa atattttaa ttccatcatt ctaacaaaat caatttttca gagtaatgat 960
tctcactgtg gagtcatttg attattaaga ccggttgga taagattaca tcctctgact 1020
ataaaatcct ggaagaaaaan ctnggaaata ttcgtctgga cattgcactt ggcaatgaat 1080
ttatgggagc tttggaatcc tgcagatata ataataataa ttaaacaaaa cactcagaga 1140
aactgccaac cctaggatga agtatattgt tactgtgctt tgggattaaa ataagtaact 1200
acagtttata gaacttttat actgatacac agacactaaa aagggaagg gtttagatga 1260
gaagctctgc tgtgcaatca agaattctag ccactcattt ctgtaggggc tgcaggagct 1320
ccctgtaaag agaggttatg gagtctgtag cttcaggtaa ccactgatcc acttcagtc 1380
actatccatg agtttttatt tccaaatata tgaaatcata tgaatttctg gtttttctctg 1440
ttggagccca aggagcaagg gcagaatgag gaacatgatg tttcttaccg acagttactc 1500
atgacgtctc catccaggac tgaggggggc atccttctcc atctaggact gggggcatcc 1560
ttctccatcc agtattgggg gtcatecttc tccatccagt attgggggtc atcctcctcc 1620

```

atccaggacc tgaggggtgt ccttttctgc gcttccttgg atggcagtct ttcccttcat 1680  
gtttatagtg atttaccatt aaatcactgt gccgtttttt cct 1723

<210> 3

<211> 876

<212> DNA

<213> Homo sapiens

<400> 3

gttccagaca gaagaaatag caagtgccga gaagctggca tcagaaaaac agaggggaga 60  
tttgtgtggc tgcagccgag ggagaccagg aagatctgca tgggtgggaag gacctgatga 120  
tacagaggtg agaaataaga aaggctgctg actttaccat ctgaggccac acatctgctg 180  
aaatggagat aattaacatc actagaaaca gcaagatgac aatataatgt ctaagtagtg 240  
acatgttttt gcacattttcc agccccttta aatatccaca cacacaggaa gcacaaaagg 300  
aagcacagag atccctggga gaaatgcccg gccgccatct tgggtcatcg atgagcctcg 360  
ccctgtgcct ggtcccgcct gtgaggggaag gacattagaa aatgaattga tgtgttcctt 420  
aaaggatggg caggaaaaca gatcctgttg tggatattta tttgaacggg attacagatt 480  
tgaaatgaag tcacaaagtg agcattacca atgagaggaa aacagacgag aaaatcttga 540  
tggcttcaca agacatgcaa caaacaaat ggaatactgt gatgacatga ggagccaag 600  
ctggggagga gataaccacg gggcagaggg tcaggattct ggccctgctg cctaaactgt 660  
gcgttcataa ccaaatcatt tcatatttct aaccctcaaa acaaagctgt tgtaatatct 720  
gatctctacg gttccttctg ggcccaacat tctccatata tccagccaca ctcattttta 780  
atatttagtt cccagatctg tactgtgacc tttctacact gtagaataac attactcatt 840  
ttgttcaaag acccttcgtg ttgcagtgtg ctcccg 876

<210> 4

<211> 2059

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (241) .. (242)

<223> a, c, g, or t

<400> 4

ggaaaatgat gttgcccaga gccacgtgat ccaggctctc actccaaact caattcaaaa 60  
gtatgttctg gaaagctggg agggagctaa ggtgggatct aacctgtcta acatcactgc 120  
ctccaaaaca ctaaaactat attaatagca aggagagttt agaagttatc agttactcca 180  
ggatacgccc gggaaaagga aacaagatca gccaaatact agaagctgga aaacagatgg 240  
nngaattggga actggcacag ccagagaaag ccaggatgtg agcggcactg ggggaggctg 300  
tggctcttca ggtttgtgag gcagagagag tcctcagaga cccaggaatt ggtgtcccca 360  
gaaactgaagg tgaaggggag gtagggctga aatcagtttg attgggtaga aaagcagaga 420  
agcagtgaga ccctcgctct ccttgtcttc caccagagg cctcaccctc ctcagcagag 480  
gtctgggagt tcattccctg gagaggggaa cagaaatctc tggaggaaaa gctgccaggc 540  
ctgggtgggg ttatggctcc caccctataa gtgactgtgc aactgcatg ctgaggggtat 600

```

aaaaacagaa gtgaggggtg aaggacctca gatgtacaca gaacacggtc ctcaggtgta 660
cacagaagtg agcacagatg ccaggagaat aggcacctgt ggcccgggtg aaggggggtca 720
ttgaagcaga gacccacgtg gggatagagc catttttggtt ctagggtgtg ttgctgagag 780
tgagtgggtc cagtggaggg aagatctttg gagatgtgga tgggggcca ggaattgaga 840
cagcacagtg ttgatggacc acccacctgg ctctgcagat caccccaaca actgcaggat 900
tgtgaaggag aagattgagc tctattacca ggttttaaac ttcgccatga tcgtgtcttc 960
tgcactcatg atatggaaag gcttgatcgt gctcacaggc agtgagagcc ccatcggtgt 1020
gggtgctgagt ggcagtatgg agccggcctt tcacagagga gacctcctgt tcctcacaaa 1080
tttccgggaa gacccaatca gagctggtga aatagttggt tttaaagttg aaggacgaga 1140
cattccaata gttcacagag taatcaaagt tcatgaaaaa gataatggag acatcaaatt 1200
tctgactaaa ggagataata atgaagttga tgatagaggc ttgtacaaag aaggccagaa 1260
ctggctggaa aagaaggacg tgggtgggaag agcaagaggg tttttaccat atgttggtat 1320
ggtcaccata ataataaatg actatccaaa attcaagtat gctcttttgg ctgtaatggg 1380
tgcatatgtg ttactaaaaac gtgaatccta aaatgagaag cagttcctgg gaccagattg 1440
aaatgaattc tgttgaaaaa gagaaaaact aatatatttg agatgttcca ttttctgtat 1500
aaaagggaac agtgtggaga tgtttttgtc ttgtccaaat aaaagattca ccagtaaaaa 1560
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1620
aaaaaaaaaa aaaaaaaaaa aaaaagtaaa aaaaataaac taaaaaagaa ggaaaaatat 1680
gaattacaga actggagttg tttaaattta cagggcgcta gtttttaagc gatctactgg 1740
aaaaacccca tgggttttctt aattaaaaac gttgtgaaac tctccgtttt ggcaaacggg 1800
tttatctgag aaacctctcc taatattctt ttccaaaact tggccactct tatgcttcat 1860
ggaaaactgc cgcccaagtt tgggctaaaa cgcggaatt tctttgaaaa ccctctcttt 1920
tttataccat gtggctcatc tgggaaattc catttaatcg cacaaatata ctggtatctc 1980
tttgggtgtt cccatttga acaagtcctt cttttacgtg gtttctaccc cgggcacctt 2040
ttttgggtg ggccccctc 2059

```

&lt;210&gt; 5

&lt;211&gt; 718

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 5

```

aattcgcggc cgcgtcgact tttttttttt ttttttgctc tgaatttatt gcgagtga 60
aacagagaaa atcctcaagt ttaagtttct gatagcagag tgtgggagtt agagcatggg 120
gagtccagag gttccagacc cccaaaggtc tctaccaggg ccatctccgt tagtggcggt 180
ggcagccccct cttgtggcct ttttcctctc tccaaggggt caccocgcac catgccgtc 240
cccctcatct atcttgcccc ctccccacag gcccatctgc gctgaactct cgccagttcc 300
gaagtgggag agggaggggg tttacttgcc cgatcggttg tgggtttgag cttatagagg 360
cagaggagta agaacctgcg atattgaaag ctaccacat ggggcttcct tgaaggagga 420
cgtggaaggc agaaagtgcg ctgctctgag cggcgcatgt aaccgaggac cttaagctgg 480
accacggggc ttggacgatt ttttaaataa ggaaatcgac ctcatcttcc tcctcctcgt 540
cctcttcccc tgaacccccca gtccgcatgc actcacactc tttggccttt tcctcagtc 600
ccgggctcct ctttggtaaa tagatttgta ggtgtctaag tcacgtccca ccctcactcc 660
ttcccaggag aggagacagg gctaggatcc caccgcgacc gcgggccata aacacttg 718

```

&lt;210&gt; 6



&lt;211&gt; 1141

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; unsure

&lt;222&gt; (52)

&lt;223&gt; a, c, g, or t

&lt;400&gt; 6

```

gggcgcggca agggcacctg tctttggaga atgagtcgct cttcagcgac tnagggtccc 60
agctctagag gagaagaaga gaagaggaga ggagagaaga gaagaagaga ggagagaaga 120
gaagaggaga ggagagaaga gatctcctcc tgctcctgtg gcagttcccg aagaagctcg 180
gtgccatcaa caggaaagga tgagactttg gattccagaa gaacactgct ggggcttccc 240
cccgagtctg ccacttgctg taacctcaga ggatccagaa cggcagctgg tccttgctgg 300
actgttcctg tccatgtgcc tggtcatggt gctggggaac ctgctcatca tccggccatg 360
agccctgact ccacctcca cactccatg tacttcttcc tctccaacct gtccttgctt 420
gacatcggtt tcacctccac cagggtcccc cagatgactg tggacatcca gtctcgagc 480
agagtcattc cctatgcagg ctgcctgact cagaagtctc tctttgccat ttttgaggc 540
acggaagaga gacatgctcc tgagtgtgat ggcctatgac cggttttag ccatctgtca 600
ccctctatat cattcagcca tcatgaacct gtgtttctgt ggcttcctag ttttgctgtc 660
tttttttttt ctcagtcttt tagactccca gctgtacaac ttgattgcct tactaatgac 720
ctgcttcaag gaggtggaaa ttcttaattt cttctgtgac ctttctcaac accccaccct 780
tgctgttgtt gacaccttca ccaatgacat agtcatgtat ttcttgctg ccatatttgg 840
ttttcttccc atttcgggga ctttttcac ttactataaa attgtttcct ccattctgag 900
ggtttcatca tcaagtggga agtataaagc cttctccacc tgtggctctc acctgtcagt 960
tgtttgctta ttttatggaa caggctttgg aggggacctc agttcagaca tgctccttta 1020
tcccagaaaa ggtgcagtgg cctcagtgat gtacacggtg gtcacctcca tgctcaaccc 1080
ctttatctac agcctgggaa acagggatat taaaagtgtc ttgcggcggc cgacggagca 1140
c                                                                 1141

```

&lt;210&gt; 7

&lt;211&gt; 1275

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 7

```

gtagcaagta agtcttcaag aaacctttgg aagagagaag gcctgtatat gaaatgccag 60
aaggaagggg caaaagggtc agaactagga ctctgggtat ggagttcaga cacctgtgaa 120
ttgggcattc tgaaccactt gtacaatgtg caaagctgat tttttcttac catagcacag 180
ggttgctttc tattacagat gtacctatct ggaaacatgt tcattcaaca gtttgagtc 240
ctgcagttat tttcaaaaca gtttttgctt ccctcttttt gtagtaaaaca cggctctttt 300
ccagtgtgct agataaagtc tggctctggg cagtaaaggg acatggctgc tgcttatttc 360
aggaacttag gggcaagtgt cttcatgcct gtggaaacag gagccagcta gtactatttc 420
cagcaagaaa tttaagagaa aggagagatt tttattatga ttttgatttc ttactacaa 480
cattgcatgt gtctggagta tagccattac actttatgaa aaaggcaaaa tggtcatttg 540
gggtgtttta ggaagtttgc caaaaggctc ctttgtcatt ataatccttc ctaagctgcc 600

```

```

atccacgggt ttaggtcatg gatatgaaaa gtgaaagggt ttagagatga agtagtgtcc 660
cctgagtgtt taccaacctg ttaatctttt tgagatgtta attttttcat atagagcccc 720
ctaaaatctt gatggctcta gatcagtcaa gcctaagaga agacgtatgt atggaaaaaa 780
acaaaaaaca aaaaaacctt gctggattgc tagtaatatc tacttcttgg aaattaatac 840
ttcatatttt ttaaaaaaat tattgatgca ttaggaatat tttttgctta gcagttacaa 900
attttaagag gcacatatat accacggaat actatgcagc cataaaaaag gatgagttca 960
tgtcctttgt agggacatgg atgaagctgg aaaccatcat tctcaacaaa ctatcgccag 1020
gacaaacaac caaacaccgc atgttctcac tcacaggtgg gaactgaaca gtgagaacac 1080
ttggacacgg gaaggggaac atcacacact ggggcctgtc gtgggggtggg gggagcgggg 1140
agggatagca ttaggagata tacctaagt aaatgatgag ttaatgggtg cagcacacca 1200
acatggcaca ggtatacata tgtaacaaac ctgcacattg tgcacatgta cactagaact 1260
taaagtataa tttaa                                     1275

```

&lt;210&gt; 8

&lt;211&gt; 934

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 8

```

ctaagcttct cgtgaagcta gatgaataga aacacggagc atctataaac tggattaaac 60
ataggggggt acaaacctcc tggatatttt taattcttag gaagaaggag aatttagcac 120
cagactcttg aaaagcagtg gcatgaaagt tagtttgata tattgtttga gtttgtattg 180
ccttaggccc tgaatcaaga ccaatggttt gctgtagctg ttggtttcaa acaggagcta 240
agagtgatgt cttccttgtg gtctgttggc tattcagtat tccagtgcga attgccaatt 300
cagttggaag aaacatagtc tagaatgtaa tgtcattttt tttttttttt ttttttttac 360
agagcaaaac tttgttggct taaaaaccag ggagtgggcc gggggtgtgg gggctcatgg 420
cctggtaatc ccagcacttg gggaggctga ggtggggcca atcactggtc gggagatcaa 480
ggccatccgg ggctaacttg ggtgaaaccc tggctcttac taaaaataca aaggatgggg 540
tgggcgcggt gggggcacac gcccatatgt ccagctact gggagaggct agagcgcggt 600
agaatcactt gaaccagga ggcggagtgt tggcagtga gccgggatca cgtcactggc 660
gacccagcc ggggcgacag agcgagactc tggctctcaa aaacaacaa aaaaaaacac 720
ccacgacggg cgagtaaaga gtggcggggc gggggcgaaa aaaagaggcg atcagtgaac 780
gcgacgccc gttagcccaa tagcgggtgt cccggtgcac cccagtttct ttgtgccacc 840
aggggggaaa ctcgcgcggt accccaata tggggagggc cgaataaatt cggaaaagcc 900
gaggcacccg ggatagagcc gggcaaggca agat                                     934

```

&lt;210&gt; 9

&lt;211&gt; 730

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 9

```

ttttgttttg gcaaaggaaa agtctcttaa taagaatttt tctagtgaac gaaggggaaa 60
atatctatgt gaactacaat gggaaatgat acaagaacag ctcagggaag gatgagaaaa 120
gatgtttctt tgcctttcaa tgctgtctaa gtgcatgtct catgaaatga aaagacagaa 180
taaagcagtt gtgcagagca tgatcttttg aatcaaatac acatccactt actaactcag 240

```

```

tgaccctgaa caaacagctt ctctgcaacc tcagtttcct catctgtaaa atgcaggtac 300
aggttgaaca tctatcccta atccaaaaat ccaaaatgct ccaaaatctg aaattttttg 360
agcactaaca tgatgctcaa aggaaatgct cattggagca tttcaaattt cagattttca 420
gattagggat gctcaaccag taagtataat gcaaatattc caaatccgaa aaaatcccaa 480
attcaaaaca ctctgggtccc aagcatctca gataaaagat aatcaacctg tagtaacaac 540
acctacttaa acctacatag acactcagta aaagtattac tattaattct gctgcctgct 600
ttttaaaatt aaatctttta gagttttgaa ctacagttta ctttttaaaa tagaaagcaa 660
ataaacactt caaattatth taatttatgt taatatttaa ccagggtaat ttgttaaaat 720
aaattgttca                                     730

```

<210> 10

<211> 413

<212> DNA

<213> Homo sapiens

<400> 10

```

ggggtgacag agcaagatct tgacccttta aaaaaaaaaa aaaagttaac tttaaacaac 60
acacaagcat tattcattgt ttaaagttaa ctttaaaca tgaataatgc ttgtgtgttg 120
ttttcacagt attatctaaa catagaaggg taaaaatcca agtatcagcc actgttttgc 180
tctatcacat ttaatatctg tgcaccaaca atagtttgta ggattatgga atcattaaac 240
atagatgcat attattccaa ggcttattat ggatgtttgt catcactgac tttccttcaa 300
tattaaaaaa gtgctaaatt ttgattggat tatgattata atttcttcta taattgacag 360
attctgtaat acaaatgggc atgtcttcga aaaaaaagaa aaaaaagtcg acg 413

```

<210> 11

<211> 714

<212> DNA

<213> Homo sapiens

<400> 11

```

caaaagaccc tgtcaacagg attaaatata cagccaccga ctcgagagaa atattttgca 60
aaacagcgta tccggcaaag aattaatatc agaatacata atgaattctc gaaaactcga 120
agtataacaa atatcgtaca ctggaaagtg ggcaaaacca ttaataaaca tttcaccaca 180
taggatatat agatggccaa agaagcatat gaaaaagatg cgcaacatca ttagctatta 240
gggaaatgca aattaaaacc accattagga tattagtaca gaatgggttaa acatcaaaaa 300
taatagtgat aacaccaa at gccaataagg aagtggagga gaaataggat cattgatata 360
ttgttttttg gaaggtaaaa tggtagagcc ctctagaaag cagtttggtta attagaaaaa 420
aaccctaaagt atgcatgcag ttctgtaaga taaagtgtct gtccaggcat gcatacaacc 480
cagcaattgc atgcctgggc gcttacctta cagaaatgaa catttataat tacattataa 540
tatgtacaca aaattcatca cagctttatt aatagaagcc aaactctctg tgggcttctc 600
acagtgtacc cattgccaga gtaaaactgca gccttgaacc attgctcagc ctccttacct 660
atgagctatg aacactgaag caggttgcac agtgaaaaaa aaaaaagtc gacc 714

```

<210> 12

<211> 1512

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; unsure

&lt;222&gt; (1289)..(1386)

&lt;223&gt; a, c, g, or t

&lt;400&gt; 12

```

ccaaaagaat gagtcatcaa aactcagtat cacttgactc aaagagcaga gttggttgcc 60
gtcattacag tgtaacaag attttaatca gtctattaac attgtatcag attctgcata 120
tgtagtacag gctacaaagg atattgagag agccctaadc aaatacatta tggatgatca 180
gttaaaccgg ctgtttaatt tggtacaaca aaatgtaaga aaaagaaatt tcccatttta 240
tattactcat attcgagcac acactaattt accagggcct ttaactagag caaatgaaca 300
agctgacttg ctagtatcat ctgcattcat ggaagcaca gaacttcatt ccttgactca 360
tgtaaagtga ataggattaa aaaataaatt tgatatcaca tggaaacaga caaaaaatat 420
tgtacaacat tgcacccagt gtcagattct acacctggcc actcaggagg caagagttaa 480
tcccagaggt ctatgtccta atgtgttatg gcaaatggat gtcatgcacg taccttcatt 540
tggaaaattg tcatgtgtcc atgtgacagt tgatacttat tcacatttca tatgggcaac 600
ctgccagaca ggagaaagta cttcccatgt taaaagacat ttattatctt gttttcctgt 660
catgggaggt ccagaaaaag ttaaaacaga caatgggcca gggtactgta gtaaagcagt 720
tcaaaaattc ttaaatcagt ggaaaattac acatacaata ggaattctct ataattccca 780
aggacaggcc ataattgaaa gaactaatag aacactcaa gctcaattgg ttaaacaaaa 840
aaaaaggaaa agacagtaag gatataacac tcccagatg caacttaatc taacactcta 900
cactttaaat tttctaaaca tacatagaaa tcagaccact acttctgcag aacattttac 960
tggtaaaaag aacagccac atgagggaaa actgatttgg tggaaagaca acaaaaacaa 1020
aacatgggaa ataggtaagg tgataatatg ggggagaggt tttgcttggt tttcaccagg 1080
agaaaatcag cttcctgttt ggataccac tagacatttg aagttctaca atgaactcat 1140
cagggatgca aatgaaagtg cctctgcaga gacagaaaac ccgcaatcga gcatcatcga 1200
ctcgcagggt gaacaaaatg gtgatatcag aagaacagat gaagttacaa tccaccaagg 1260
aaactgcaca tgtggagagt caggagagann nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1320
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1380
nnnnnnnagg agacaaagag ataaaagggt cgagtgcaga ggtgaggaga aagactgaaa 1440
acgatgagaa acagcaacta agacacaaag gaggtgggag actgcccggg tgccacagca 1500
cccacaccgt cc

```

1512

&lt;210&gt; 13

&lt;211&gt; 1826

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 13

```

actgtcagac ctggggagcc ggcctccagc agcgggagcg gcgggagcga gcacgacccc 60
actctcctgc ggccgagggg ggagcagcgc gagcccgccg cgctgagccg gccgggggag 120
gggagatgag ttgcggcccc gcggcagcgc cccaggatgg ggagggagcg gcggcactgc 180
cctcgagaa tgccgctccg gtgaagtagg cgccgagcgc cgtccgcctc cccaagccg 240
ttccgcaccg cggccgctca gcctctgcca tggccggctc tggcgctgag aagcgctca 300

```

```

aatctatgct aaggaaggat gatgcgccgc tgtttttaa tgacaccagc gcctttgact 360
tctcggatga ggcgggggac gaggggcttt ctcggttcaa caaacttoga gttgtggtgg 420
ccgatgacgg ttccgaagcc ccggaaggc ctgttaacgg ggcgcaccog accctccagg 480
ccgacgatga ttcttactg gaccaagact tacctttgac caacagtcag ctgagtttga 540
aggtggactc ctgtgacaac tgcagcaaac agagagagat actgaagcag agaaaggatga 600
aagccagggt gaccattgct gccgttctgt acttgctttt catgattgga gaacttgtag 660
gtggatacat tgcaaatagc ctagcaatca tgacagatgc acttcatatg ttaactgacc 720
taagcgccat catactcacc ctgcttgctt tgtggctatc atcaaaatca ccaaccaaaa 780
gattcacctt tggatttcat cgcttagagg ttttgtcagc tatgattagt gtgctgttgg 840
tgtatatact tatgggattc ctcttatatg aagctgtgca aagaactatc catatgaact 900
atgaaataaa tggagatata atgctcatca ccgcagctgt tggagttgca gttaatgtaa 960
taatgggggt tctgttgaa cagtctggtc accgtcactc ccattcccac tccctgcctt 1020
caaattcccc taccagagggt tctgggtgtg aacgtaacca tgggcaggat agcctggcag 1080
tgagagctgc atttgtacat gctttgggag atctggtaca gagtgttgggt gtgctaatag 1140
ctgcatacat catacgaatc aagccagaat acaagattgc tgaccccatc tgtacatacg 1200
tattttcatt acttgtggct tttacaacat ttcgaatcat atgggataca gtagttataa 1260
tactagaagg tgtgccaagc catttgaaatg tagactatat caaagaagcc ttgatgaaaa 1320
tagaagatgt atattcagtc gaagatttaa atatctggtc tctcacttca ggaaaatcta 1380
ctgccatagt tcacatacag ctaattcctg gaagttcatc taaatgggag gaagtacagt 1440
ccaaagcaaa ccatttatta ttgaacacat ttggcatgta tagatgtact attcagcttc 1500
agagttacag gcaagaagtg gacagaactt gtgcaaattg tcagagttct agtccctaata 1560
tttatgtatt gttttagcat tgctgaatc actttattta tcctgcagtc acagacttga 1620
gagcaataaa tgcaaaccta aatgagaaaa tggaaatccct gacagctgtg tccgtatcaa 1680
gcatcagctc ctcaaacagt tgccccagcc tgacagtgtc agtctctgtt taatggtaaa 1740
aggagacttt gccataattt tcagatgaag atgtttccca aacactgttt acagaatgag 1800
atgtgactcc tacagatacc tcatag 1826

```

<210> 14

<211> 1454

<212> DNA

<213> Homo sapiens

<400> 14

```

gattggagca cgaaccctgt tgtgaactgc ctatccgaag gatctagggt gtgtgcttcg 60
tatgagaatc taatgccaga tgatctatca ttgtctcact ttgccccag ataagacat 120
ctagttgcag aaaaataagc tcagagcttc cactgattct acattatgga tatgtgccgc 180
cgaagcaagc acaaagccct acttttacac atgcctagt atgcttcacg gacaaggctt 240
ggctctgttg agtccaacta acctacctga gattctgaga tttctcttca atggcttcct 300
gtgagctaga gtttgaaaat atcttaaaat cttgagctag agatggaagt agcttggacg 360
attttcatta tcatgtaaat cgggtcactc aaggggcaa ccacagctgg gagccactgc 420
tcaggggaag gttcatatgg gactttctac tgcccaagggt tctatacagg atataaagg 480
gcctcacagt atagatctgg tagcaaagaa gaagaaacaa aactgatct ctttctgcca 540
cccctctgac cttttggaac tcctctgacc ctttaggaca agcctacct atactgtcta 600
gagaaaagac caacaacggc ctcaaaggat ctcttaccat gaaggtctca gctaattctt 660
ggctaagatg tgggttccac attaggttct gaatatgggg ggaagggtca atttgtcat 720
tttgtgtgtg gataaagtca ggatgcccag gggccagagc agggggctgc tgctttggga 780
acaatggctg agcatataac cataggtatg ggaacaaaa acatcaaagt cactgtatca 840

```

```

attgccatga agactcgagg gacctgaatc taccgattca tottaaggca gcaggaccag 900
tttgagtggc aacaatgcag cagcagaatc aatggaaaca acagaatgat tgcaatgtcc 960
ttttttttct cctccttctg acttgataaa agggaccgtc ttccttggat ttagtgaacc 1020
cctttggttc ctgaaaaatt caaggagtat ctaggacata gtccccagaa gacagtacaa 1080
gactttctga taaactggac atttcaagac ccaaataact aatcagaaaa atcaaagatg 1140
tgatactatt ttttatccca tgcataagtg ctacacttgg atcaaagtga caatgttggg 1200
atctctatgg ataaagggtct taaaagtcct gagataaaga atcctgcacc cactggtaac 1260
ttctaataga aacaaacaac acgcatgaaa aaaaaaacat ggtcgacggc taggccctaa 1320
gaaagattaa aaaccgcgtt ggggggacggc accagagtca gaaaagtggc acgggtgcc 1380
aggaaccggg aagaacaggg gtggtctcgc aaaaaacggg ggggcacggg tttgacaccg 1440
aaggataggg acgg                                     1454

```

&lt;210&gt; 15

&lt;211&gt; 1361

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 15

```

gaaaatgaag ggagaagaaa attcataaaa aataaagctt ggtcaagcag aaggccaaaa 60
aagaaaagaa gaaaaaaatg aatgagaaaa ggaaaaggga agagggttct gatattgaag 120
atgaggacat ggaagaactt cttaatgaca caagactctt gaaaaaactt aagaaaggca 180
aaataactga agaagaattt gagaagggtt tggtgacaac tggcaaaaga acaatcaaga 240
cagtggattt agggatctca gatttggaag atgactgcta attccagtgc cacagatgaa 300
cccacaagga catagctgtt ccctaacttg gtggatggct ccagtttctt tttaacgaaa 360
atcacaactt caggagacat ctgaaaagaa tgatgtctct gaaagctgtc ctttcagatg 420
agggagaaat gaaggatttc acacttcaga atattttact aaaaacattc cagtcttggc 480
cgggtgcggg ggctcctgcc tataatccca gcactttggg aggctgaggc aggaggatca 540
cttgagccca ggagttcaag accagcctgg gaacacagcg agaccctctc attaaaaaca 600
acaaaacaaa acaattccag tcttgagtag gtctaacaga agaaaatgta aaattatttg 660
agtgtaaata atagatgtca gtatttatca tgatgggtca catatagaca tatgtacata 720
ttatatatat atatatatat atatatatat atatatatat atataagctc 780
ttttttctga ggctatttta tagttatttt taaacataaa gatacagaag tcttcttgac 840
ttctgatttt caaaaccatt cctcagtatc ttcaggcatt tgacctcctg aatgtgcttg 900
gccctgggct tcagttatcc tttgatgtcc tgcaggggtg gctaattgtc tgggggtttt 960
ctgtgttaat agtcacagta ttgttttatt ggtgaatagc tgaaaaacag agggattaa 1020
tcatattccg ggaaagagaa ttatagtttt tatgcctcct gttgaataaa tgggtgtcctg 1080
attgcctggg tottaagtgt gtaaagttag ctggacagtg ttttccccta gatggcctgg 1140
tcagcttgca gtattgatgc aacaccacat cagcgtcttt catagagttt atttgaagaa 1200
atgctgaatt tattcccaag tataatttta aaaagctgtt taggacccaa acatatttta 1260
acatctctta cacatacaga atttcagttt acaaatattc cagaaggcat tttcttaagc 1320
agtgaaccat gcagaatact gttgctaaac atcgtttctg c                                     1361

```

&lt;210&gt; 16

&lt;211&gt; 1418

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 16

```

atgctgtcac tgetccacgc atcaacgctg gcagtccttg gggctctgtg tgtatatggt 60
gcaggtcacc tagagcaacc tcaaatttcc agtactaaaa cgctgtcaaa aacagcccg 120
ctggaatgtg tgggtgtctgg aataacaatt tctgcaacat ctgtatatgt gtatcgagag 180
agacctgggtg aagtcataca gttcctgggtg tccatttcat atgacggcac tgtcagaaag 240
gaatccggca ttccgtcagg caaatttgag gtggatagga tacctgaaac gtctactacc 300
actctcacca ttcacaatgt agagaaacag gacatagcta cctactactg tgccttgtgg 360
gagggtgcggc tagccaacca agagttgggc aaaaaaatca aggtatttgg tcccgaaca 420
aagcttatca ttacagataa acaacttgat gcagatgttt cccccaagcc cactattttt 480
cttccttcaa ttgctgaaac aaagctccag aaggctggaa catacctttg tcttcttgag 540
aaatttttcc ctgatgttat taagatacat tggcaagaaa agaagagcaa cagattctg 600
ggatcccagg aggggaacac catgaagact aacgacacat acatgaaatt tagctgggta 660
acggtgccag aaaagtcact ggacaaagaa cacagatgta tcgtcagaca tgagaataat 720
aaaaacggag ttgatcaaga aattatcttt cctccaataa agacagatgt catcacaatg 780
gatcccaaag acaattgttc aaaagatgca aatgatacac tactgctgca gtcacaaaac 840
acctctgcat attacatgta cctcctcctg ctctcaaga gtgtgggtcta ttttgccatc 900
atcacctgct gtctgcttag aagaacggct ttctgctgca atggagagaa atcataacag 960
acggtggcac aaggaggcca tcttttctc aatcggttat tgtccctaga agcgtcttct 1020
gaagatctag ttgggctttc tttctgggtt tgggccatc cagttctcat gtgttgtcta 1080
ttcctatcat tattggatta ccggttttca aaccagtggg cacacagaga acctcactct 1140
gtaataacaa tgaggaatag ccacggcgat ctccagcacc aatctctcca tgttttccac 1200
agctcttcca gccaaaccaa atagcgctg ctatagtgtg gacatctcgc gcttctagcc 1260
ttgtccctct cttagtgttc tttaatcaga taactgcctg gaagcctttc attttacacg 1320
ccctgaagca gtcttctctg ctagttgaat tatgtggtgt gtctttccgt aataagcaac 1380
atcaaattta aaaaaatgaa aactcaacaa caaaacaa 1418

```

&lt;210&gt; 17

&lt;211&gt; 11461

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; unsure

&lt;222&gt; (4908)

&lt;223&gt; a, c, g, or t

&lt;220&gt;

&lt;221&gt; unsure

&lt;222&gt; (5097)

&lt;223&gt; a, c, g, or t

&lt;400&gt; 17

```

gccgtgctt ctttcacact ttagttggga gctgcgcgcc gcgctcagtt actggagagc 60
tgccgcgcgc ccgcgcctc ccgcacgctt gcacgcgggc ccggcttcgg ggttttgggt 120
tcttactcca agcggcgggg agggggggga gccccggaca cactgtgggg aggaggagga 180
agaagaggag gagggaggaa gaaaaaagac gaggaggaca ggggcggggg gcgggaggct 240

```

```

tgccaccttc agcccccccg cgaacgcccc aggtgcacac atcttgacca actcagcagc 300
aaggtggatt ttctttgtgt ttaaagaaaa aaaatgtccc tgtgtctgta gagatgattt 360
gcagttcagc cgggctgaag ctgaccgaat gagactatgg gctgtgcctc cgccaagcat 420
gttgccactg ttcaaaatga agaggaagcc cagaaaggga aaaactacca gaacggagat 480
gtgtttggcg atgagtatag gatcaaacca gtggaagagg tcaaatatcat gaaaaatggg 540
gcagaagaag agcagaaaaat agcagccagg aaccaagaaa acttggaata aagtgccagc 600
tcaaatgtaa gacttaaaac taataaagag gttccgggat tagttcatca acccagagca 660
aacatgcaca tctctgaaag ccaacaagaa ttcttcagaa tgctggatga aaaaattgaa 720
aagggtcggg attactgttc ggaagaagag gatatcacat agcaccaatt ttaccactca 780
aaccaggagc tactactgtg taaatagggt acacccagat tgaaatcttt gcaaaggctg 840
gttctattca gcgaacagca ctatagcaaa agaagatcgt tccatattgt acgccccatt 900
aaattacagt gtttcttaat gaacttgcaa aggaatattg ctaaaaacaa acaaaaaaaa 960
ctgttatcga actttctttg ttgctgctag ttaaaacttg ttgcaacttt tcaacttctct 1020
tgtgtccagg tatgcagcaa aattctgcaa ttccacctta aagatactgt tgggttttaca 1080
gatgctctcc aacctatttt ctataagatg aggtagtggg gaactcagat aacaaacttc 1140
tcttctaaac tggttctgct tctaagacaa gcatctcctg ccctctctcc ttctctccca 1200
tctctgcac gcagtctaga gatggactga gccttgcttc tcaactggcag tgttgagctt 1260
tgagatggg atggttgcta tgccaagcct tgtttctctg ctcaagaaga gtagagaagc 1320
tattatcaat taaaagcatg ctgtgatgtg actcctggaa gtgacgtagg agtgagtggc 1380
aggttgtttg atttaaatag tatcttaate aagaattaag cttgcaacat tggctttgct 1440
cagatgcaga tggaagtgtg atcacaatca ttttgaatcc ctcttctca cttttttttc 1500
taagaaaata aacattttac tgtttttatg gatccttgct ttctccatt catccagctc 1560
agtgttttaa gatgatcctg ggtgcagaag ttgagccctc ctttgcatg acactgataa 1620
ttagcctata gggctcccta cccttccatt aagaatctac caagcattag caaggctgaa 1680
agtgtctaa gaggtgaggg gacatcctat gactttttag gaaggcctga aaccaccttg 1740
ttacctttca ttttgtagc aaataaacca tcctattttg taactctccc ccttcaaaat 1800
gctacatgag ccttgccact tcctttttct cttacttcca gcacactaga catagcaaaa 1860
gtgtttgcct actcaaaaac ataatacttt tatgtgatg atggtatttg gagatgtgaa 1920
agccaaaagc ccctggcagt ggtgggaat gttgactgag tgttcagcag agtttatttt 1980
tccatactat atgaagagaa tgatctcttc tcaaagacag aagtgatatt ttttaacaaat 2040
attgtcacia gtaaatagca atcaaaagga gaaaataact tttgtatttt tttaatgtgt 2100
ttgatagctt tgacgagggg tctctttggt actttcaggg gagggcatcc tattaaatgc 2160
cacgccagca gtccgggtct gggtttgctc caaaaatca caggagcact gtatgttctt 2220
ctcttttgga gttgtgactt tgaaggcct caatattagc cacactgccg cctgcagaag 2280
gtggagagtt aagatgttct atgtcaattt gctcttgccg aaaagatgag cctcgatttt 2340
aaaatctatc cacatccaac tgatggcacc attgatgtgc aaataatgag attccctatc 2400
tccttttaga cctgggacgg caaaagggaa gggaaggaaa cttagcagag tgctattgac 2460
tatagattca catattagca acaaaatccc gtaattcttt tggccaacag cagctatttt 2520
ggggagcagc tgtggctgtt acataaatag agatgcagcc aaaatttttag gccttttatc 2580
ctgcttctag cagaaaaatg cagggagagt caagtgtct agggtttcag gttgcctccc 2640
ctcatatggt ttttgccaa gtgactaaaa cagttttcca caactgtaaa caactgcta 2700
agccccacct caaacttggt cactggggac tttgcttacc gttctgtggg tgaccttttc 2760
cgggatttct tgttcttacc aagcaagaat taagcacatg ctaaacgtct tccatttgac 2820
ttctctactc ggtgtctcag acagtgtctt cccagaaaac caccacctc taccacaaaga 2880
tgaaacatgc tcatgtcatt tttctcatgg tcacatttaa cagttttgac atgttatact 2940
tgcgcataga tccaagcgtt tcttggaac ctgacttttg agtgtttaat aaagccggaa 3000
gtggtgttgc cctgaaccag cagattttca cctgggttct ggctccggtg ttttaacctg 3060
gatacatctt tgatgtgcga aagtgagttc atcttcagac acatttggtg catccagaaa 3120

```



```

tagatccaag aaatggggtg gttgagtggg tccgcacgaa atgcttgatt atgtcagcaa 3180
caccacaacac tgtctgtttt ccatttgttg gttttaatca taaaattgtc aagtgattcg 3240
tgtttgtact ttattttttt gtgccttctg aaaggatcta aaacaaaaat attttgcctt 3300
tttttcccca cgtgtatctg aacattaagc agattggctc agacacaatg aaaaggataa 3360
tccaatgtac gtgctgggtg actctgctag ttgttatctc tgtagggctc aggaagctgg 3420
aaggaggaag ggagggtaag tggcctgggtg agtggaggta gaaaaatgat gagaaatgaa 3480
ctgagagcat taagcagaga gggttgatag gctggccgtg tccgggggtga aattggaaat 3540
ccagctgcct agtggccagt ggggtgggca agactgtcaa cgagatttac agctggctta 3600
cacatgcctt atgtcctctg agttgtagag ttgtaaaagt tcagcagtgt gtgcacagct 3660
ttcttttggg tggcagagat tcaggatcat ggagtactgc tctcataatt gaagacgtgt 3720
ttgttattgg cagagaacct taaaaaaggc ctttacctca gcgatgcttc ctagecccg 3780
gcttgacagag aacacagagt ggtgttgggt tctatttagg gacaaagaag gtataaagtc 3840
cagagatgag aaaactgggt cagccctcag aaactgcagc agccacgcac acagaagcct 3900
gctggaagac aggtctctct ccgtccacag tgcccatcat ctgagcctgg gctgggatga 3960
ctcaacttag caaagacgga cccaggagga gtgctgggtc ttcagtcttt gtactggccc 4020
catcctctcc tcaactgtaat gtgaggaagc acctctgtgt cagggtcac ctgggcatcc 4080
aaagcgcca cgccacaat ccgacagccc ccaggagcag gtccagggat gatgcagccc 4140
ccttcttggg cccatgtgat gtcacctgc tttgttatct tctataact ttatcctgct 4200
ataactttat cctcttccca gcctcatccc tgttttctg ttagggaag actcttccat 4260
aagcctgcta aaaaacagag atgatactc ttacaaactt tacctcatag cctgtgaagc 4320
aggttggcat gtggattaca agtcctgctt tgacactggg caagaattta agattgttct 4380
atctctacta gtcatagaaa agaaacattg ttaaacatgt tgagttttaa aggaagaaat 4440
attttcaaat tcttaatacca agaaaatact gagttggaat cttagacttc gggactctga 4500
cacgttcttt atgaaaggca aaataattgg tatctaaagt tctctccttc ctgcctcccc 4560
taaagaaaaa ggatattaga ttgcacacta taattttaca taagatctgc cttccacact 4620
tccctgctgg aaggcattct cagagcttta tgtcttcgta cctctcagag attggacttt 4680
ttcttgttta aaaccccaac caaaaaaat aataaggcat gattgggtggg gagggaatgt 4740
gtatttaggg gcataataaa aagggtggctc gaagcaggaa ctttggcctc atggtgtcat 4800
gggtggatgc ctggactcca gtgtgcctgt gaggggctgg gttaggcagt cggctgtcac 4860
actcacatgt gcctgcaata aaccttttgg aatttcatga acgaggtnc tgaattgcc 4920
ttttgccaat gaatggatgt atttttccaa ggggggaata gtatccttga ctttggcagt 4980
cacctttttg tatgtctcta garaggggtc aaaaaaacta tggtaaagat gaggcttatg 5040
agtgaatacc tctgggacaa accttaggac tcacaagcta tgccatgttt ttcagganac 5100
tctgtaccta tctggaatcc aatcctggga gaagaggaaa aaggagctaa tatttgcct 5160
ttatactcac tctgtgccag atactgtgsc aggcatttat aattgtttgt gtcaccccat 5220
gatatccgtt aaattaatct attacccccg tgttctaaat aatagatctt aagtgtggaa 5280
tgcatctatc caacataaat gccatgttg aaagaaggaa agatgtcatt caagtatttt 5340
tcaaattctt tttattatga ctatgccct cgcaacactg tgaacacaac ccctgggggc 5400
atctgccttc cagaatctct ctctggcttc tcaccagctt ggtttcctca tggggagtgt 5460
tttatttggc ctccccatc tgagctgcac acacaccagg gggaggccac tggctaacag 5520
taggacttca gtgccctgag gaaaaggctt tgggaattta ggcacacgtt tctctcctt 5580
gaatccttct agcatctgga aaaggaacct ggtatttctg cagttagata ccagatgcaa 5640
ccagaacagg ctttcgagtg ctgtgatttc tttcctgggt tccaagtct tgttgtttca 5700
tcacaaactg tatcttttta aggttaaaag tcttgacct catgggggtc tgggacaatc 5760
cgatctccaa gcatggagga aaggcaatgc ctggaccact gacttgcatt gaaatcctt 5820
cttgtgggct aggttttgat gtctctttt catctttgga ctggggatct gcatcttcca 5880
ggtccattta agcactgaaa ctagatgcaa atctcttctg agaccttaca tgttttagat 5940
agtcatgtaa tgacttggat agacatttaa ataacttgtt ccaaggctcg aagagcccaa 6000

```

gaactcctca gagttcctct tcttgcttct tgaatcacat cctctaaaga tgactcatgt 6060  
 ctccatagca actgttaaag gtgctgccta gacgggaccc gctcccaoct ttacttactt 6120  
 tgctcgaagg aaccaaaca atgtccttgt gtaaaggggc tgtcatatcc agttttcctt 6180  
 tgaaatctgg cccccaagat cctgcttctt tctaacctgg cagctgtcat gtcccttcaa 6240  
 gatgatgtgg gaaatggccc ttacaacttg ggaaccacag aaattgctgt atttcgggaa 6300  
 gattcacctc taaactgaag gcttcattct gatagtgtct gccctctcta ccctgatttc 6360  
 gcccttcttt gcttccattt ttagcccaag gctttgaatt tgattgagta agacttagag 6420  
 gcagtataaa gaacaccata aacttaggca gaggtccctt aggggtctcta gaggtgaaaa 6480  
 taattctaca gccttagggg gacctcttgg cattgactct aaaggagag aatagcccct 6540  
 gtgtcctggc atttcagtct agacctcaa ggactgttct ctcttgacag gcaagcaagc 6600  
 aaagaaagtt ttgcaataga tttcaagcca gtttttccat tcaaaccaag atgcaaattc 6660  
 ataaaattac tcttttctct gaatagatcc aggcagctgc cttattagaa ctttagatto 6720  
 ggatctattt tcttaacaca catacacata caccgcata catacatata cagagagata 6780  
 cgtggagaaa ggaaatttac tctatcattg caatacttca agaaagagct gtattttgcc 6840  
 tttctgtaat ctccaagata gtgtctagga aagtaatagt ataactatag ggataccgaa 6900  
 acaggaaaaa ccagccatca ctcttgagaa agtttgagtt cgactcacat gggagaatcg 6960  
 aggtctgcta ctgctcttgc tttgtgcccc atctgtgcct ggatgccta ctacayctgc 7020  
 ttgactctgc tgggctgcta gccggggtgt tgtggtgac atcctttcct ggcttacac 7080  
 acataataga cacatcccta acggcgtgtg cctgggtccag ccacatacag ccaccacatg 7140  
 tgtcacacac tgtccccctc atccagtgtg gacttgactg gcatttcagc agctccactg 7200  
 ggatgctcta accccagtgt gtggagttgg ggtcccttca tctaggttga cccaggtata 7260  
 gcatttttag cattgccttt ccagtcttga tgattcattc attgaactca tttatttctg 7320  
 gagcccctgg tactctccag gcactgcgct aattgccagc aaagcacaac tgaactaaat 7380  
 ccaccttcaa ggaacctagc cataacgagg gaggcagcat ggaagtacc tacaggggaa 7440  
 agtctgaggt gctgtgggag catctcaccg tggcagccag cccagttttg gcaatcaggg 7500  
 gcttctgaa agaggtgaca tcaaagcccg gatgtgtcag aggactgagg gagagtgtta 7560  
 ctaaaggact ttcaggctga gaggatagca caggactcag ccaaaggag ggccagtgtg 7620  
 gactgtccag ggccagcctg cagtacagag gctggagctt ggactttag agggagagag 7680  
 aagagcaagg gacgtggacg gggcagtgag ccaggctagc cacagagggg tccccgggct 7740  
 ttgtctggga ttcaggggagc ataaataaga gctttagggt gtgctgtgtc ctctgcagcc 7800  
 cactgctgag gtctccaga caggttaagg gtggtcaca tcagggcccg ggtttccctg 7860  
 ctactgcgg cagtgcagg gtgcttctg agatgattca tcccagggtg tcctctgtcc 7920  
 cttaaccagc ccaactcct ctctctctgc caaaagctat ttgaattcaa ggactttaac 7980  
 ctgggcccga tctggttttg agacaaagg gacagctctg ggtcagcatg accttcttta 8040  
 gagccactaa ggcgaaaaat accgtttggg accaggctgg cctagacca gggatgagaa 8100  
 tgcaccctaa aataaatata cgggaagcag cagagggtt ccctgtctag tgtgtgatcc 8160  
 taactaaagg cagctctctt ggacagcctt cccctggatt aggtcacata cacctgggtg 8220  
 ccaagcctct gctgggtccc aaatacacac ccgagtcctg ccaaagaaag gagattttta 8280  
 aaaagcacag acaaattgta tgcaagtggg aaatacccat aggcctagac agctgtggag 8340  
 ggaagacctc gtgggtacct ggaggtgcc agagctggga gctctgcagg tatgagtcag 8400  
 ggaaggctca gagacaagca gaatctctct atggagacaa cttgcagtgc cttttagggt 8460  
 ttccaaataa cctcgaggtt cagagcattg ggttttttcc tccccctccc acccccagaa 8520  
 aaataattag aaaaatgttt aggagaaagg aaaagaatta gatgcattcag aataccagct 8580  
 ataagccaac actgtttcca gaaactcaag aaaaagctca aacagaagac agttcccctg 8640  
 agaggctgga ggcgttggtg ctgaaggcaa ttttcttagc taaggggcac tgggccttgc 8700  
 tgcaccttg ggctgacct ttttgcaaaa ccccccccc tgccctctct gcataactcaa 8760  
 cagcaacgcc agctttctgg acccttgga agatgttagc tcaaaccacc actttttcca 8820  
 gatcttctc ttgctcttca ctgaggaatt tgtaattctg aggttagcga tgccactcg 8880

```

gatattccgc aggccaggt gtttagatta gaatttgtcc agcggtaatc ctgatgctgg 8940
aaaccaacaa acatttggcc tcatattcac ccatttaaaa actagagccc ctggcaggtc 9000
cccttagggc catgtgttca tggaatataa gccaaagttg ccttaggctt gttcatggaa 9060
tataagccaa gtttacctct cccattttc tgccctggcc cacttcccac tcacctccac 9120
ctcattgcc a ggaagggatc aaaatgcctc catgccagtt gttaatggct acatatttgc 9180
ccttcccaag ggtatttgca ttttatttag gaacatggcc ttatatcca ggaaaaacta 9240
gcatcaagat tacgaggcat cacctctcaa tcaggtctgg gaggtatctt ggggcattgc 9300
tcttctgaac acctgcagag gcttcctcag gtgagtgtgg gagcccgaa ggggtggctc 9360
cctaaccact ctgcctgcac atgaattctc caaagcagtg ggccccatc tgtttcaatt 9420
acacatgcct gtcagcaaaa acttcgtgag atgcactctc tctgtgtgtt tattaattta 9480
tttaaagcat atatcccttt acttttgtag tactatatta ggcacattat aaaaagtata 9540
cagcatagaa actttaaatg aataagacac aaaatattat aaacagagg tctggcattt 9600
tctctctgaa ctctgaggg ggaccttggg cacctcctgg tatgtgcaca cccactttg 9660
aacacccttg ctttagtgca aatcagcacg cctaaatagc atcccaaacc actgtgcgac 9720
at ttggcctg cagaataaga aagtcctaag gcagaatccc caggaaactc aaatctggga 9780
gatgaggaaa gaaaactcac tcacaaacaa cataaatgta aataattcaa aaccctaaag 9840
gagagtgtgc cctagacagt agtggctaca ggtgctaagc aaaagtcagg tacttgagc 9900
agagccatgc gattgattgt tcatcttccc tctctgggtc tccagaaaag gacactggca 9960
atatccccag ctctctcctg attggaatc tttgaaccct cgaagtgtc cagcagtact 10020
acccccacg tagcagtggg cccatgtcgc ttgaatatta tttttgattt gaccaccaag 10080
atgtaatgat gattctattc cattttgaaa agggctctctg actggaactt ggggagcttc 10140
cttagtctca tgccgaagag gagtcaagct taggagacct ctggtctgct tcttgggtgct 10200
ccctgaatgc ttgttagtga cgtcagcttc cagtttactg gtaacagctg gactgctgcc 10260
ggaggccagt aacctggagt tctcttgggtg gtcctctcct gacactacca cgttcctttg 10320
cctatgctgg ctcttgctgt cccttcgcca agaaaactct cccatttctc ctccccactg 10380
ggaaccaagc ctctctgag ttccagttaa tattgagcac tcacatgaag tctcctcacc 10440
gtatgacct a catatgcc ttctgttctg ccaaagaaca ttccctttag gaggggcccg 10500
gtgtgggtgg caaggtgggg taggaatccc tgacaaggat gatccaggga gaattctgg 10560
gagtaagacc tggctctcaa gtgcaacttg gtcttcaaga gccaaagttg catgctgagg 10620
aaggttgacc ccagagcaga gcctaggcca ggtgaggagc aggtagccct tttggacttg 10680
ccacagctct ttccacaagg actcagagtc cctcagggga gagcaccccc cagagttttg 10740
agcaacattg actgactggc tgagatgacc tgggacagca ggtccactc tcaggcaacc 10800
gagccacagg gagtgtttga gaacaacttg tccaccctgc atgccactga gttgggcacc 10860
agcaacaaa gctgaacaga ctaaagcctg cctctgaaga acttgagat gataggggtg 10920
aggtttgcaa acacgtactc caacgggctc tccctgcccc tccagagtgc tctgttcata 10980
ccgctgatgc tgagatgggc agccactgag cagcatgac atgagtaacc attccgtcct 11040
cttggcctgg gtttccttcc caagggctgt ggtttggggc atgttttgac ttgaaacagt 11100
cagaccatg gcggtatttg gagtgtgaag cccctctctg gtgtcactcc ctcccaggtc 11160
tttaaaccca ggattttctg attccagggc catgattctt tcatcgtgct ctctaggcag 11220
aggggtgctg cctgcacata tcacctcac tgctgagcca ctctgtctcc ccacagccat 11280
tgatgatgga ggtgaaatag cacctcctac aacagcgtct atgtcctgag gctcttctcc 11340
ccaaatttgg ccctcttctc tgctcctatc gctgtcccca cagtgcaggc ccacatcatc 11400
tcttgctgag tatcgaagta gtttccaaac acgtcctccc tctgttctct tttcctctac 11460
a
11461

```

&lt;210&gt; 18

&lt;211&gt; 277

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 18

```

ctactactac taaattcgcg gcccgtcgac gtaacagcag ttaggaggct gattatgata 60
atggaacttc ttaactttaa ttttccttag actaaaagtt ctctgaactc cttcctcaga 120
aagttgtatt gaaattccac acattaggag cctaattgtc cccaggccaa cctacctggg 180
agaaataagt gtgctgtgac cctctatgtc atggataaat tgatatctca ttcttagcat 240
aaaaatgaca aacaaaagaa tggtcgacgc ggccgcg          277

```

&lt;210&gt; 19

&lt;211&gt; 2125

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 19

```

ccatcttcgt ggccgcctgt ccgctcgcgc cgctcttcgc cctgctcaac aactgggtgg 60
agatccgctt ggacgcgcgc aagttcgtct gcgagtaccg gcgcccgggtg gccgagcgca 120
cccaggacat cggcatctgg ttccacatcc tggcgggcct cacgcacctg gcggtcatca 180
gcaacgcctt cctcctggcc ttctcgtccg acttcctgcc gcgcgcctac taccggtgga 240
cccgcgcccc cgacctgcgc ggcttcctca acttcacgct ggcgcgagcc cctcctcct 300
tcgccgcgcg gcacaaccgc acgtgcaggt atcgggcttt cgggatgac gatggacatt 360
attcccagac ctactggaat cttcttgcca tccgcctggc cttcgtcatt gtgtttgagg 420
tagccgaggc acctgctggt tctcccatcc atggcatgag gccccgaccc tgtgctttgc 480
ctaattcgag cacgtggtga ggggtcggtg ccgtcacttc ctgctgtgtc atcttggtca 540
aatcagagct cttctctgca cctgcgtttt ccctgcctgg cctcatccct gggttgtggt 600
gtggacattg tgggtgtctc cacaggagcc ccagggccac gaaagctggg gtggcctctg 660
ccccttctgg ggttcctttt cctgcacagc tgctttctga ctccaccac agctgggagc 720
agggtccgga gccccggcct gcctggccct gtgaaggcca ctctgggcgt ttgggtgggc 780
gtgagtgcct tcctctgtc ccagcatgtg gttttctcgc ttggccgcct cctggacctc 840
ctggtgcctg acatcccaga gtctgtggag atcaaagtga agcgggagta ctacctggct 900
aagcaggcac tggctgagaa tgaggttctt tttggaacga acggaacaaa ggatgagcag 960
cccgagggtc cagagctcag ctcccactgg acaccttca cggttcccaa ggccagccag 1020
ctgcagcagt gacgcctgga aggacatctg gtggtcctta ggggagtggc ccctcctgag 1080
ccctgcgagc agcgtccttt tcctcttccc tcaggcagcg gctgtgtgaa ccgtggctg 1140
gctgttgtgc ctcatctctg ggcacattgc ctgcttcccc ccagcgccgg cttctctctt 1200
cagagcgctt gtcactccat ccccggcagg gagggaccgt cagctcacia ggccctcttt 1260
gtttcctgct ccagacata agcccaaggg gccctgcac ccaagggacc ctgtccctcg 1320
gtggcctccc caggccccctg gacacgacag ttctcctcag gcagggtggc tttgtggtcc 1380
tcgccgcccc tggccacatc gccctctcct cttacacctg gtgaccttcg aatgtttcag 1440
agcgcagggc cgttctccct cgtgtcctct ggaccacccc gcccttcct gccctgtttg 1500
cgcagggaca tcaccacat gccccagctc tcggaccctg cagctctgtg tcccaggcca 1560
cagcaaaggt ctgttgaaac cctccctcca ttcccagtta tctgggtcct ctggattctt 1620
ctgtttcttg aatcaggctc tgctttcccc ctagccacta caggcagcct ctgacagtgc 1680
cgctttactt gcattctgca gcaattacat gtgtcctttt gatccttgcc caacttcctt 1740
ccctctcccc gctcctggcc cctggcccag ggccctctt gctgttttta cctctgttcc 1800
ttggggccta gtaccagca agcaccacaa tgggggaggt tttgggatga gaggaggaaa 1860

```

cgtgtataacc tgtaacatct ggtggctctt cccccagaag tttgtgttca tacataattg 1920  
 ttttccacgc tggatcataa tgtgacgtgc agttctgccc tgtgctgggg agccacatga 1980  
 agcttccctt ggctaacttg ctaccccgca gcaatcccag tgtggccgtc tgcttgctaa 2040  
 aaaatggatc tgtgctcatc tgtattgatg tccttgaggc tctacaagtg gaacttaagt 2100  
 gtcaaaaaga atatgtgggt ttttag 2125

<210> 20

<211> 200

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (180)

<223> a, c, g, or t

<400> 20

gctcgagaat ccaacttgaa aagtgtctga tgaaactatg cttgaaagaa atcaggctgc 60  
 agctaaaaat atataatgat gataaggagg aatagaaaag taagttaaac aagtgttctc 120  
 ttttttctct tctggctgcc aaagtggatt tagaaaaatg aattactctc ccaactactn 180  
 cgaggaggctg aggcaggaga 200

<210> 21

<211> 256

<212> DNA

<213> Homo sapiens

<400> 21

atgaaaataa tgctactctg cccactaatt ttttttcttg gaaacacaat tttttaaaat 60  
 gaaaaatcta atttaattcc ttctttcccc atgtgcataa aaaatctaatt ttatattcat 120  
 gttaatatgt aataagtgtg ttatttgtaa atgaataaac aaacatctaa aatttttctc 180  
 agttttaatt tctaatacag tatacatcaa tagatatgac tcatataaac aaaacttctt 240  
 tggggctctga tatgggt 256

<210> 22

<211> 585

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (70)

<223> a, c, g, or t

<220>

&lt;221&gt; unsure

&lt;222&gt; (163)

&lt;223&gt; a, c, g, or t

&lt;400&gt; 22

```

gtttgctttt gctttctctg cccggataat gggaactctg aaatcgaaag caattaaatg 60
tgatgttggn aaacagttat aaacttgtgt cttcttggag agtaggctat gtagtctttc 120
tggcagcagt agcaaggaaa taatgaggtg aaagcgccca ggnaagaagg ccccatgaac 180
actgcaaaga cagtttgctt gttgtaaatg gaggtttctc ctggcgggaa ggctgtgacc 240
ctgctgagct gtggagaaac agcttgataa ccagggtctt cctggggcac cctttctggt 300
gcacttcgga agaccccgga ctgatgatag ggctctcttg atttcaccta aagcaagact 360
gtaacctctg ggcagcctga tgccaggctt ggaggcaaac cctcagaaaa cagctcccct 420
cagtgagggg atgggagcct gtttgaatta acgtggagaa ttctccaaga acccgagac 480
tgggagggcat tctcacggcg agtcagtga gtctcttat gtttaccgca attaaggtga 540
gcatgagagc agcctgcctc tcttctgttc tccagccat ggaag 585

```

&lt;210&gt; 23

&lt;211&gt; 99

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 23

```

Leu Leu Met Ala Phe Asp Asp Leu Thr Ala Lys Leu Lys Ala Val Ser
  1              5              10              15

```

```

Ser Lys Tyr Val Tyr Leu Thr Lys Lys Glu Val Ile Gln Gln Glu Leu
      20              25              30

```

```

Leu Ser Met Lys Lys Ile Gln Gln Glu Tyr Lys Lys Leu Glu Glu Asp
      35              40              45

```

```

Lys Lys Met Leu Glu Lys Val Ile Leu Asn Leu Lys Lys His Met Glu
      50              55              60

```

```

Ser Thr His Trp Ala Arg His Ile Gly Glu Val Gln Lys Tyr Lys Leu
      65              70              75              80

```

```

Asp Leu Glu Glu Arg Ala Ile Gln Ala Ile Glu Lys Leu Val Glu Ile
      85              90              95

```

Pro Leu Lys

&lt;210&gt; 24

&lt;211&gt; 184

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 24

Pro Pro Gly Ser Ala Asp His Pro Asn Asn Cys Arg Ile Val Lys Arg  
 1 5 10 15

Lys Ile Glu Leu Tyr Tyr Gln Val Leu Asn Phe Ala Met Ile Val Ser  
 20 25 30

Ser Ala Leu Met Ile Trp Lys Gly Leu Ile Val Leu Thr Gly Ser Glu  
 35 40 45

Ser Pro Ile Val Val Val Leu Ser Gly Ser Met Glu Pro Ala Phe His  
 50 55 60

Arg Gly Asp Leu Leu Phe Leu Thr Asn Phe Arg Glu Asp Pro Ile Arg  
 65 70 75 80

Ala Gly Glu Ile Val Val Phe Lys Val Glu Gly Arg Asp Ile Pro Ile  
 85 90 95

Val His Arg Val Ile Lys Val His Glu Lys Asp Asn Gly Asp Ile Lys  
 100 105 110

Phe Leu Thr Lys Gly Asp Asn Asn Glu Val Asp Asp Arg Gly Leu Tyr  
 115 120 125

Lys Glu Gly Gln Asn Trp Leu Glu Lys Lys Asp Val Val Gly Arg Ala  
 130 135 140

Arg Gly Phe Leu Pro Tyr Val Gly Met Val Thr Ile Ile Met Asn Asp  
 145 150 155 160

Tyr Pro Lys Phe Lys Tyr Ala Leu Leu Ala Val Met Gly Ala Tyr Val  
 165 170 175

Leu Leu Lys Arg Glu Ser Glu Ser  
 180

&lt;210&gt; 25

&lt;211&gt; 75

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 25

His Ile Tyr Thr Thr Glu Tyr Tyr Ala Ala Ile Lys Lys Asp Glu Phe  
 1 5 10 15

Met Ser Phe Val Gly Thr Trp Met Lys Leu Glu Thr Ile Ile Leu Asn  
20 25 30

Lys Leu Ser Pro Gly Gln Asn Asn Gln Thr Pro His Val Leu Thr His  
35 40 45

Arg Trp Glu Leu Asn Ser Glu Asn Thr Trp Thr Arg Glu Gly Glu His  
50 55 60

His Thr Leu Gly Pro Val Val Gly Trp Gly Glu  
65 70 75

<210> 26

<211> 422

<212> PRT

<213> Homo sapiens

<400> 26

Pro Lys Glu Ala Val Ile Lys Thr Gln Tyr His Leu Thr Gln Arg Ala  
1 5 10 15

Glu Leu Val Ala Val Ile Thr Val Leu Gln Asp Phe Asn Gln Ser Ile  
20 25 30

Asn Ile Val Ser Asp Ser Ala Tyr Val Val Gln Ala Thr Lys Asp Ile  
35 40 45

Glu Arg Ala Leu Ile Lys Tyr Ile Met Asp Asp Gln Leu Asn Pro Leu  
50 55 60

Phe Asn Leu Leu Gln Gln Asn Val Arg Lys Arg Asn Phe Pro Phe Tyr  
65 70 75 80

Ile Thr His Ile Arg Ala His Thr Asn Leu Pro Gly Pro Leu Thr Arg  
85 90 95

Ala Asn Glu Gln Ala Asp Leu Leu Val Ser Ser Ala Phe Met Glu Ala  
100 105 110

Gln Glu Leu His Ala Leu Thr His Val Asn Ala Ile Gly Leu Lys Asn  
115 120 125

Lys Phe Asp Ile Thr Trp Lys Gln Thr Lys Asn Ile Val Gln His Cys  
130 135 140

Thr Gln Cys Gln Ile Leu His Leu Ala Thr Gln Glu Ala Arg Val Asn



145	150	155	160
Pro Arg Gly Leu Cys	Pro Asn Val Leu Trp Gln Met Asp Val Met His		
165	170	175	
Val Pro Ser Phe Gly Lys Leu Ser Phe Val His Val Thr Val Asp Thr			
180	185	190	
Tyr Ser His Phe Ile Trp Ala Thr Cys Gln Thr Gly Glu Ser Thr Ser			
195	200	205	
His Val Lys Arg His Leu Leu Ser Cys Phe Pro Val Met Gly Val Pro			
210	215	220	
Glu Lys Val Lys Thr Asp Asn Gly Pro Gly Tyr Cys Ser Lys Ala Val			
225	230	235	240
Gln Lys Phe Leu Asn Gln Trp Lys Ile Thr His Thr Ile Gly Ile Leu			
245	250	255	
Tyr Asn Ser Gln Gly Gln Ala Ile Ile Glu Arg Thr Asn Arg Thr Leu			
260	265	270	
Lys Ala Gln Leu Val Lys Gln Lys Lys Gly Lys Asp Ser Lys Asp Ile			
275	280	285	
Thr Thr Pro Gln Met Gln Leu Asn Leu Thr Leu Tyr Thr Leu Asn Phe			
290	295	300	
Leu Asn Ile His Arg Asn Gln Thr Thr Thr Ser Ala Glu His Phe Thr			
305	310	315	320
Gly Lys Lys Asn Ser Pro His Glu Gly Lys Leu Ile Trp Trp Lys Asp			
325	330	335	
Asn Lys Asn Lys Thr Trp Glu Ile Gly Lys Val Ile Ile Trp Gly Arg			
340	345	350	
Gly Phe Ala Cys Val Ser Pro Gly Glu Asn Gln Leu Pro Val Trp Ile			
355	360	365	
Pro Thr Arg His Leu Lys Phe Tyr Asn Glu Leu Ile Arg Asp Ala Asn			
370	375	380	
Glu Ser Ala Ser Ala Glu Thr Glu Asn Pro Gln Ser Ser Ile Ile Asp			
385	390	395	400
Ser Gln Gly Glu Gln Asn Gly Asp Ile Arg Arg Thr Asp Glu Val Thr			

405                                      410                                      415  
 Ile His Gln Ile His Gln  
 420  
  
 <210> 27  
 <211> 438  
 <212> PRT  
 <213> Homo sapiens  
  
 <400> 27  
 Met Ala Gly Ser Gly Ala Trp Lys Arg Leu Lys Ser Met Leu Arg Lys  
     1                                      5                                      10                                      15  
  
 Asp Asp Ala Pro Leu Phe Leu Asn Asp Thr Ser Ala Phe Asp Phe Ser  
                                     20                                      25                                      30  
  
 Asp Glu Ala Gly Asp Glu Gly Leu Ser Arg Phe Asn Lys Leu Arg Val  
                                     35                                      40                                      45  
  
 Val Val Ala Asp Asp Gly Ser Glu Ala Pro Glu Arg Pro Val Asn Gly  
                                     50                                      55                                      60  
  
 Ala His Pro Thr Leu Gln Ala Asp Asp Asp Ser Leu Leu Asp Gln Asp  
     65                                      70                                      75                                      80  
  
 Leu Pro Leu Thr Asn Ser Gln Leu Ser Leu Lys Val Asp Ser Cys Asp  
                                     85                                      90                                      95  
  
 Asn Cys Ser Lys Gln Arg Glu Ile Leu Lys Gln Arg Lys Val Lys Ala  
                                     100                                      105                                      110  
  
 Arg Leu Thr Ile Ala Ala Val Leu Tyr Leu Leu Phe Met Ile Gly Glu  
                                     115                                      120                                      125  
  
 Leu Val Gly Gly Tyr Ile Ala Asn Ser Leu Ala Ile Met Thr Asp Ala  
                                     130                                      135                                      140  
  
 Leu His Met Leu Thr Asp Leu Ser Ala Ile Ile Leu Thr Leu Leu Ala  
     145                                      150                                      155                                      160  
  
 Leu Trp Leu Ser Ser Lys Ser Pro Thr Lys Arg Phe Thr Phe Gly Phe  
                                     165                                      170                                      175  
  
 His Arg Leu Glu Val Leu Ser Ala Met Ile Ser Val Leu Leu Val Tyr  
                                     180                                      185                                      190

Ile Leu Met Gly Phe Leu Leu Tyr Glu Ala Val Gln Arg Thr Ile His  
 195 200 205

Met Asn Tyr Glu Ile Asn Gly Asp Ile Met Leu Ile Thr Ala Ala Val  
 210 215 220

Gly Val Ala Val Asn Val Ile Met Gly Phe Leu Leu Asn Gln Ser Gly  
 225 230 235 240

His Arg His Ser His Ser His Ser Leu Pro Ser Asn Ser Pro Thr Arg  
 245 250 255

Gly Ser Gly Cys Glu Arg Asn His Gly Gln Asp Ser Leu Ala Val Arg  
 260 265 270

Ala Ala Phe Val His Ala Leu Gly Asp Leu Val Gln Ser Val Gly Val  
 275 280 285

Leu Ile Ala Ala Tyr Ile Ile Arg Phe Lys Pro Glu Tyr Lys Ile Ala  
 290 295 300

Asp Pro Ile Cys Thr Tyr Val Phe Ser Leu Leu Val Ala Phe Thr Thr  
 305 310 315 320

Phe Arg Ile Ile Trp Asp Thr Val Val Ile Ile Leu Glu Gly Val Pro  
 325 330 335

Ser His Leu Asn Val Asp Tyr Ile Lys Glu Ala Leu Met Lys Ile Glu  
 340 345 350

Asp Val Tyr Ser Val Glu Asp Leu Asn Ile Trp Ser Leu Thr Ser Gly  
 355 360 365

Lys Ser Thr Ala Ile Val His Ile Gln Leu Ile Pro Gly Ser Ser Ser  
 370 375 380

Lys Trp Glu Glu Val Gln Ser Lys Ala Asn His Leu Leu Leu Asn Thr  
 385 390 395 400

Phe Gly Met Tyr Arg Cys Thr Ile Gln Leu Gln Ser Tyr Arg Gln Glu  
 405 410 415

Val Asp Arg Thr Cys Ala Asn Cys Gln Ser Ser Ser Pro Cys Ala Asn  
 420 425 430

Cys Gln Ser Ser Ser Pro  
 435

<210> 28  
 <211> 93  
 <212> PRT  
 <213> Homo sapiens

<400> 28  
 Glu Asn Glu Gly Arg Arg Lys Phe Ile Lys Asn Lys Ala Trp Ser Lys  
   1                  5                  10                  15  
 Gln Lys Ala Lys Lys Glu Lys Lys Lys Lys Met Asn Glu Lys Arg Lys  
                   20                  25                  30  
 Arg Glu Glu Gly Ser Asp Ile Glu Asp Glu Asp Met Glu Glu Leu Leu  
           35                  40                  45  
 Asn Asp Thr Arg Leu Leu Lys Lys Leu Lys Lys Gly Lys Ile Thr Glu  
       50                  55                  60  
 Glu Glu Phe Glu Lys Gly Leu Leu Thr Thr Gly Lys Arg Thr Ile Lys  
   65                  70                  75                  80  
 Thr Val Asp Leu Gly Ile Ser Asp Leu Glu Asp Asp Cys  
                   85                  90

<210> 29  
 <211> 318  
 <212> PRT  
 <213> Homo sapiens

<400> 29  
 Met Leu Ser Leu Leu His Ala Ser Thr Leu Ala Val Leu Gly Ala Leu  
   1                  5                  10                  15  
 Cys Val Tyr Gly Ala Gly His Leu Glu Gln Pro Gln Ile Ser Ser Thr  
           20                  25                  30  
 Lys Thr Leu Ser Lys Thr Ala Arg Leu Glu Cys Val Val Ser Gly Ile  
       35                  40                  45  
 Thr Ile Ser Ala Thr Ser Val Tyr Trp Tyr Arg Glu Arg Pro Gly Glu  
       50                  55                  60  
 Val Ile Gln Phe Leu Val Ser Ile Ser Tyr Asp Gly Thr Val Arg Lys  
   65                  70                  75                  80  
 Glu Ser Gly Ile Pro Ser Gly Lys Phe Glu Val Asp Arg Ile Pro Glu

[illegible]

```
<210> 30
<211> 122
<212> PRT
```

&lt;213&gt; Homo sapiens

&lt;400&gt; 30

Met Gly Cys Ala Ser Ala Lys His Val Ala Thr Val Gln Asn Glu Glu  
 1 5 10 15

Glu Ala Gln Lys Gly Lys Asn Tyr Gln Asn Gly Asp Val Phe Gly Asp  
 20 25 30

Glu Tyr Arg Ile Lys Pro Val Glu Glu Val Lys Tyr Met Lys Asn Gly  
 35 40 45

Ala Glu Glu Glu Gln Lys Ile Ala Ala Arg Asn Gln Glu Asn Leu Glu  
 50 55 60

Lys Ser Ala Ser Ser Asn Val Arg Leu Lys Thr Asn Lys Glu Val Pro  
 65 70 75 80

Gly Leu Val His Gln Pro Arg Ala Asn Met His Ile Ser Glu Ser Gln  
 85 90 95

Gln Glu Phe Phe Arg Met Leu Asp Glu Lys Ile Glu Lys Gly Arg Asp  
 100 105 110

Tyr Cys Ser Glu Glu Glu Asp Ile Thr Thr  
 115 120

&lt;210&gt; 31

&lt;211&gt; 139

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 31

Ile Phe Val Ala Ala Cys Pro Leu Ala Pro Leu Phe Ala Leu Leu Asn  
 1 5 10 15

Asn Trp Val Glu Ile Arg Leu Asp Ala Arg Lys Phe Val Cys Glu Tyr  
 20 25 30

Arg Arg Pro Val Ala Glu Arg Thr Gln Asp Ile Gly Ile Trp Phe His  
 35 40 45

Ile Leu Ala Gly Leu Thr His Leu Ala Val Ile Ser Asn Ala Phe Leu  
 50 55 60

Leu Ala Phe Ser Ser Asp Phe Leu Pro Arg Ala Tyr Tyr Arg Trp Thr  
 65 70 75 80

Arg Ala His Asp Leu Arg Gly Phe Leu Asn Phe Thr Leu Ala Arg Ala  
85 90 95

Pro Ser Ser Phe Ala Ala Ala His Asn Arg Thr Cys Arg Tyr Arg Ala  
100 105 110

Phe Arg Asp Asp Asp Gly His Tyr Ser Gln Thr Tyr Trp Asn Leu Leu  
115 120 125

Ala Ile Arg Leu Ala Phe Val Ile Val Phe Glu  
130 135

<210> 32

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

<400> 32

ggaggcagaa gaggaagtca ga

22

<210> 33

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

<400> 33

gccatccatg tttctcagtt cc

22

<210> 34

<211> 31

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

<400> 34

atgtgctgtg aaatctacta ccgtttgctg g

31

<210> 35

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

<400> 35

ggtggcatgt gtttttactt tca

23

<210> 36

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

<400> 36

aacgggtctt aataatcaaa tgactc

26

<210> 37

<211> 35

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

<400> 37

aatttccatc attctaacaa aatcaatttt tcaga

35

<210> 38

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

<400> 38

ggtcgtatgt atttccaggt gag

23



<210> 39

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

<400> 39

ttgtcatctt gctgtttcta gtgat

25

<210> 40

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

<400> 40

aggctgctga cttaccatc tgaggc

26

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/29386

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C07H 21/02, 21/04; C12N 15/85; C12P 21/06

US CL : 536/23.2; 435/325, 69.1

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.2; 435/325, 69.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
MPSRCH, DIALOG, WEST

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	Database Genbank, Accession No. AI420333, NCI-CGAP. 'National Cancer Institute, Cancer Genome Anatomy Project (CCAP), Tumor Gene Index'. 28 March 1999. See entire document.	1,3,5-12,14-15,18-19,86 ----- 21-22,24-35,58

☒ Further documents are listed in the continuation of Box C.

See patent family annex.

Special categories of cited documents:		"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier application or patent published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

13 January 2002 (13.01.2002)

Date of mailing of the international search report

07 FEB 2002

Name and mailing address of the ISA/US

Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703)305-3230

Authorized officer

ANTHONY CAPUTA

Telephone No. 703-308-0916

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/29386

## C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P — Y,P	Database Genbank, Accession No. AX037575, VINALS, Y.D. 'Compositions and methods for the therapy and diagnosis of prostate cancer'. WO 00/60084-A (SMITHKLINE BEECHAM BIOLOG ET AL), 12 October 2000. See entire document.	1,3,5-12,14-15,18-19,86 ----- 21-22,24-35,58

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/29386

## Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claim Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claim Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:  
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-35,58,86,98-99

Remark on Protest

☐

The additional search fees were accompanied by the applicant's protest.

☐

No protest accompanied the payment of additional search fees.

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/29386

### BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This International Search Authority has found 366 inventions claimed in the International Application covered by the claims indicated below:

Groups 1-22, claim(s) 1-35, 58, 86, and 98-99, drawn to an isolated nucleic acid molecule of SEQ ID NO: 1-22, a vector comprising said nucleic acid molecule, a host cell comprising said vector, a method for producing a polypeptide encoded by said nucleic acid molecule, an antisense of said nucleic acid molecule, and probes. Each of a nucleic acid molecule of SEQ ID NO: 1-22 is a distinct invention, i.e. group, and not a species.

Groups 23-45, claim(s) 36-48, 55-57, and 85, drawn to a polypeptide encoded by a nucleic acid molecule of SEQ ID NO: 1-22.

Groups 46-67, claim(s) 49-54, 100-114, drawn to an antibody that specifically binds to a polypeptide encoded by a nucleic acid molecule of SEQ ID NO: 1-22.

Groups 68-89, claims 59-66, drawn to a method for determining the presence of a PSNA, by detecting a nucleic acid molecule of SEQ ID NO: 1-22.

Groups 90-101, claims 67-71, drawn to a method for determining the presence of a PSP, by detecting a polypeptide encoded by a nucleic acid molecule of SEQ ID NO: 1-22, using a specific antibody.

Groups 102-123, claims 72-78, 80, drawn to a method for detecting cancer, by detecting a nucleic acid molecule of SEQ ID NO: 1-22.

Groups 124-145, claims 72-77, 79-80, 94-95 drawn to a method for detecting cancer, by detecting a polypeptide encoded by a nucleic acid molecule of SEQ ID NO: 1-22.

Groups 146-167, claim 81, drawn to a method for identifying a genetic lesion from a nucleic acid molecule of SEQ ID NO: 1-22.

Groups 168-189, claim 82, drawn to a method for diagnosis of a non-cancerous prostate disorder, by detecting a nucleic acid molecule of SEQ ID NO: 1-22.

Groups 190-211, claim 82, drawn to a method for diagnosis of a non-cancerous prostate disorder, by detecting polypeptide encoded by a nucleic acid molecule of SEQ ID NO: 1-22.

Groups 212-234, claim 83, drawn to a method for producing an engineered prostate cell from a pluripotent cell.

Groups 235-256, claims 84, 87, drawn to a method for treating cancer, comprising administering a prostate specific nucleic acid, or an antisense molecule of a nucleic acid molecule of SEQ ID NO: 1-22.

Groups 257-278, claim 84, drawn to a method for treating cancer, comprising administering a prostate specific protein.

Groups 279-300, claims 88-90, drawn to a method for treating cancer, comprising administering an antibody that specifically binds to a polypeptide encoded by a nucleic acid molecule of SEQ ID NO: 1-22.

Groups 301-322, claims 91-92, drawn to a method for identifying an agent that binds to a polypeptide encoded by a nucleic acid molecule of SEQ ID NO: 1-22.

Groups 323-344, claim 93, drawn to a method for identifying an agent that downregulates a nucleic acid molecule of SEQ ID NO: 1-22.

Group 345, claims 96-97, drawn to a computer readable means for storing a nucleic acid molecule of SEQ ID NO: 1-22, or a polypeptide encoded by a nucleic acid molecule of SEQ ID NO: 1-22.

and it considers that the International Application does not comply with the requirements of unity of invention (Rules 13.1, 13.2 and 13.3) for the reasons indicated below:

The inventions listed as Groups 1-366 do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

An international stage application shall relate to one invention only or to a group of invention so linked as to form a single general inventive concept. If multiple products, processes of manufacture or uses are claimed, the first invention of the category first mentioned in the claims of the application will be considered as the main invention in the claims, see PCT article 17(3) (a) and 1.476 (c), 37 C.F.R. 1.475(d). Group I will be the main invention. After that, all other products and methods will be broken out as separate groups (see 37 CFR 1.475 (d)).

Group 1, drawn to a nucleic acid of SEQ ID NO:1, and a method of producing a polypeptide encoded by said nucleic acid, forms a single inventive concept.

Groups 2-67, 345 are additional products, which are structurally different from SEQ ID NO:1.

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/29386

Groups 68-344 are additional methods, which have different method objectives, steps and reagents used than those of the method of group 1.